



Guida al rilascio di Amazon EMR

Amazon EMR



Amazon EMR: Guida al rilascio di Amazon EMR

Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e il trade dress di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, secondo qualsiasi modalità che possa causare confusione tra i clienti o secondo qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Informazioni sui rilasci di Amazon EMR	1
Versioni di rilascio di Amazon EMR 6.x	2
Versioni dell'applicazione nei rilasci di Amazon EMR 6.x	3
emr-6.14.0	3
emr-6.13.0	32
emr-6.12.0	60
emr-6.11.1	95
emr-6.11.0	124
emr-6.10.1	159
emr-6.10.0	189
emr-6.9.1	230
emr-6.9.0	260
emr-6.8.1	305
emr-6.8.0	334
emr-6.7.0	378
emr-6.6.0	422
emr-6.5.0	468
emr-6.4.0	495
emr-6.3.1	528
emr-6.3.0	554
emr-6.2.1	586
emr-6.2.0	612
emr-6.1.1	644
emr-6.1.0	663
emr-6.0.1	689
emr-6.0.0	706
Versioni del rilascio di Amazon EMR 5.x	727
Versioni dell'applicazione nei rilasci di Amazon EMR 5.x	730
emr-5.36.1	730
emr-5.36.0	760
emr-5.35.0	798
emr-5.34.0	824
emr-5.33.1	848
emr-5.33.0	877

emr-5.32.1	899
emr-5.32.0	922
emr-5.31.1	949
emr-5.31.0	968
emr-5.30.2	991
emr-5.30.1	1009
emr-5.30.0	1031
emr-5.29.0	1053
emr-5.28.1	1071
emr-5.28.0	1089
emr-5.27.1	1108
emr-5.27.0	1125
emr-5.26.0	1143
emr-5.25.0	1162
emr-5.24.1	1181
emr-5.24.0	1198
emr-5.23.1	1216
emr-5.23.0	1232
emr-5.22.0	1250
emr-5.21.2	1268
emr-5.21.1	1284
emr-5.21.0	1301
emr-5.20.1	1319
emr-5.20.0	1335
emr-5.19.1	1354
emr-5.19.0	1370
emr-5.18.1	1388
emr-5.18.0	1403
emr-5.17.2	1420
emr-5.17.1	1436
emr-5.17.0	1451
emr-5.16.1	1468
emr-5.16.0	1484
emr-5.15.1	1500
emr-5.15.0	1515
emr-5.14.2	1531

emr-5.14.1	1547
emr-5.14.0	1562
emr-5.13.1	1579
emr-5.13.0	1594
emr-5.12.3	1609
emr-5.12.2	1624
emr-5.12.1	1639
emr-5.12.0	1653
emr-5.11.4	1670
emr-5.11.3	1684
emr-5.11.2	1699
emr-5.11.1	1714
emr-5.11.0	1728
emr-5.10.1	1744
emr-5.10.0	1758
emr-5.9.1	1774
emr-5.9.0	1788
emr-5.8.3	1804
emr-5.8.2	1818
emr-5.8.1	1832
emr-5.8.0	1847
emr-5.7.1	1862
emr-5.7.0	1876
emr-5.6.1	1891
emr-5.6.0	1905
emr-5.5.4	1920
emr-5.5.3	1934
emr-5.5.2	1948
emr-5.5.1	1962
emr-5.5.0	1976
emr-5.4.1	1991
emr-5.4.0	2005
emr-5.3.2	2020
emr-5.3.1	2034
emr-5.3.0	2048
emr-5.2.3	2062

emr-5.2.2	2076
emr-5.2.1	2090
emr-5.2.0	2105
emr-5.1.1	2119
emr-5.1.0	2133
emr-5.0.3	2147
emr-5.0.0	2160
Versioni del rilascio di Amazon EMR 4.x	2175
Versioni dell'applicazione nei rilasci di Amazon EMR 4.x	2176
Differenze nella versione del rilascio	2177
emr-4.9.6	2219
emr-4.9.5	2233
emr-4.9.4	2246
emr-4.9.3	2260
emr-4.9.2	2274
emr-4.9.1	2287
emr-4.8.5	2301
emr-4.8.4	2315
emr-4.8.3	2328
emr-4.8.2	2343
emr-4.8.0	2356
emr-4.7.4	2370
emr-4.7.2	2383
emr-4.7.1	2396
emr-4.7.0	2409
emr-4.6.0	2423
emr-4.5.0	2436
emr-4.4.0	2447
emr-4.3.0	2459
emr-4.2.0	2469
emr-4.1.0	2479
emr-4.0.0	2487
Versioni AMI 2.x e 3.x	2494
Creazione di un cluster	2495
Installazione di applicazioni	2498
Personalizzazione delle configurazioni	2498

Hive	2506
HBase	2516
Pig	2529
Spark	2535
S3DistCp	2538
Novità	2542
Approccio per mitigare il CVE-2021-44228	2542
Soluzione per l'operazione di bootstrap di Amazon EMR relativa al CVE-2021-44228 e al CVE-2021-45046 di Log4j	2543
Domande frequenti	2551
Amazon EMR 6.14.0	2553
Amazon EMR 5.36.1	2556
Compatibilità con SigV4	2562
Archive (Archivia)	2563
Versione 6.14.0	2553
Versione 6.13.0	2566
Versione 6.12.0	2568
Versione 6.11.1	2576
Versione 6.11.0	2578
Versione 6.10.0	2588
Rilascio 6.9.0	2603
Rilascio 6.8.0	2620
Rilascio 6.7.0	2639
Rilascio 6.6.0	2658
Rilascio 5.35.0	2679
Rilascio 5.34.0	2683
Rilascio 6.5.0	2685
Rilascio 6.4.0	2688
Rilascio 5.32.0	2696
Rilascio 6.2.0	2701
Rilascio 5.31.0	2709
Rilascio 6.1.0	2714
Rilascio 6.0.0	2721
Rilascio 5.30.1	2726
Rilascio 5.30.0	2731
Rilascio 5.29.0	2736

Rilascio 5.28.1	2737
Rilascio 5.28.0	2739
Release 5.27.0	2741
Rilascio 5.26.0	2743
Rilascio 5.25.0	2746
Release 5.24.1	2749
Rilascio 5.24.0	2750
Rilascio 5.23.0	2752
Rilascio 5.22.0	2754
Rilascio 5.21.1	2758
Rilascio 5.21.0	2759
Rilascio 5.20.0	2761
Rilascio 5.19.0	2765
Rilascio 5.18.0	2767
Rilascio 5.17.1	2768
Rilascio 5.17.0	2768
Rilascio 5.16.0	2769
Rilascio 5.15.0	2771
Rilascio 5.14.1	2772
Rilascio 5.14.0	2772
Rilascio 5.13.0	2774
Rilascio 5.12.2	2775
Rilascio 5.12.1	2775
Rilascio 5.12.0	2775
Rilascio 5.11.3	2777
Rilascio 5.11.2	2777
Rilascio 5.11.1	2778
Rilascio 5.11.0	2778
Rilascio 5.10.0	2779
Rilascio 5.9.0	2781
Rilascio 5.8.2	2783
Rilascio 5.8.1	2784
Rilascio 5.8.0	2784
Rilascio 5.7.0	2786
Rilascio 5.6.0	2787
Rilascio 5.5.3	2788

Rilascio 5.5.2	2788
Rilascio 5.5.1	2788
Rilascio 5.5.0	2789
Rilascio 5.4.0	2790
Rilascio 5.3.1	2791
Rilascio 5.3.0	2791
Rilascio 5.2.2	2792
Rilascio 5.2.1	2792
Rilascio 5.2.0	2793
Rilascio 5.1.0	2794
Rilascio 5.0.3	2795
Rilascio 5.0.0	2795
Rilascio 4.9.5	2797
Rilascio 4.9.4	2797
Rilascio 4.9.3	2798
Rilascio 4.9.2	2798
Rilascio 4.9.1	2798
Rilascio 4.8.4	2799
Rilascio 4.8.3	2799
Rilascio 4.8.2	2800
Rilascio 4.8.0	2801
Rilascio 4.7.2	2801
Rilascio 4.7.1	2802
Rilascio 4.7.0	2803
Rilascio 4.6.0	2804
Rilascio 4.5.0	2806
Rilascio 4.4.0	2807
Rilascio 4.3.0	2808
Rilascio 4.2.0	2810
Configurazione delle applicazioni	2812
Configurazione delle applicazioni durante la creazione di un cluster	2815
Fornitura di una configurazione nella console quando si crea un cluster	2815
Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster	2815
Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster	2816
Riconfigurazione di un gruppo di istanze in un cluster in esecuzione	2817
Considerazioni sulla riconfigurazione di un gruppo di istanze	2818

Riconfigurazione di un gruppo di istanze nella console	2821
Riconfigurazione di un gruppo di istanze utilizzando la CLI	2822
Riconfigurazione di un gruppo di istanze utilizzando Java SDK	2827
Risoluzione dei problemi	2828
Come memorizzare dati di configurazione sensibili in AWS Secrets Manager	2831
Creazione di un segreto	2831
Come concedere ad Amazon EMR l'accesso per recuperare il segreto	2831
Come utilizzare il segreto in una classificazione di configurazione	2832
Come aggiornare il valore del segreto	2833
Configurazione delle applicazioni per l'uso di una macchina virtuale Java specifica	2833
Considerazioni	2833
Sovrascrittura della JVM	2836
Porte di servizio	2838
Utenti delle applicazioni	2840
Verifica delle dipendenze mediante il repository degli artifact	2841
File system EMR (EMRFS)	2844
Visualizzazione coerente	2846
Attivazione della visualizzazione coerente	2850
Monitoraggio di oggetti in Amazon S3 con la visualizzazione coerente di EMRFS	2852
Logica di ripetizione dei tentativi	2853
Metadati della visualizzazione coerente EMRFS	2855
Configurazione di notifiche di coerenza per CloudWatch e Amazon SQS	2858
Configurazione della visualizzazione coerente	2860
Riferimento ai comandi CLI di EMRFS	2865
Autorizzazione di accesso ai dati di EMRFS in Amazon S3	2876
Creazione di un provider di credenziali personalizzato per i dati di EMRFS in Amazon S3 ..	2877
Gestione dell'endpoint AWS Security Token Service predefinito	2878
Configurazione della crittografia di Amazon S3 mediante le proprietà di EMRFS	2879
Utilizzo di AWS KMS keys per la crittografia EMRFS	2881
Crittografia lato server di Amazon S3	2882
File crittografato lato client Amazon S3	2884
Delta Lake	2892
Introduzione	2892
Utilizzo dei cluster Delta Lake	2893
Delta Lake con Flink	2893
Delta Lake con Trino	2897

Delta Lake con Spark	2899
Delta Lake con Spark e Glue	2904
Considerazioni	2905
Cronologia	2905
Flink	2907
Creazione di un cluster con Flink	2909
Configurazione di Flink	2910
Hive e Glue	2910
File di configurazione	2912
Nodi primari multipli	2913
Dimensioni del processo di memoria	2914
Dimensione del file di output del log	2915
Java 11	2916
Processi Flink	2921
Avvio di un'applicazione YARN di Flink come fase su un cluster di lunga durata	2921
Invio di lavoro a un'applicazione Flink esistente su un cluster di lunga durata	2923
Invio di un processo transitorio Flink	2924
Shell Scala di Flink	2926
Interfaccia utente Flink	2927
Flink con Zeppelin	2928
Introduzione	2928
Prerequisiti	2929
Configurazione di Zeppelin-Flink su un cluster EMR	2929
Esecuzione dei processi Flink con Zeppelin-Flink su un cluster EMR	2931
Cronologia dei rilasci di Flink	2936
Ganglia	2972
Creazione di un cluster con Ganglia	2973
Visualizzazione dei parametri di Ganglia	2975
Parametri Hadoop e Spark in Ganglia	2976
Cronologia delle versioni di Ganglia	2976
Hadoop	3020
Configura Hadoop	3021
Configurazione attività	3022
Impostazioni di configurazione Daemon Hadoop	3362
Configurazione HDFS	3604
Crittografia trasparente in HDFS su Amazon EMR	3605

Configurazione della crittografia trasparente HDFS	3606
Considerazioni per la crittografia trasparente HDFS	3609
Server di gestione delle chiavi Hadoop	3609
Crittografia trasparente HDFS su cluster EMR con più nodi master	3613
Creazione ed esecuzione di un'applicazione Hadoop	3615
Creazione di file binari con Amazon EMR	3615
Elaborazione di dati con lo streaming	3617
Elaborazione di dati con un JAR personalizzato	3623
Attiva il riconoscimento non uniforme degli accessi alla memoria per i container YARN	3627
Cronologia delle versioni di Hadoop	3629
Note sul rilascio di Hadoop per versione	3687
HBase	3691
Creazione di un cluster con HBase	3694
Creazione di un cluster con HBase utilizzando la console	3694
Creazione di un cluster con HBase utilizzando la AWS CLI	3695
HBase su Amazon S3 (modalità di archiviazione Amazon S3)	3696
Abilitare HBase su Amazon S3	3697
Utilizzo della replica di lettura al cluster	3698
Tracciamento HFile persistente	3700
Considerazioni operative	3702
Utilizzo della shell HBase	3706
Creare una tabella	3707
Inserimento di un valore	3707
Ottenimento di un valore	3707
Eliminazione di una tabella	3707
Accesso a tabelle HBase con Hive	3708
Utilizzo di snapshot HBase	3709
Creazione di uno snapshot utilizzando una tabella	3710
Eliminazione di uno snapshot	3710
Visualizzazione delle informazioni relative allo snapshot	3710
Esportazione di uno snapshot in Amazon S3	3710
Importazione di uno snapshot da Amazon S3	3711
Ripristino di una tabella da snapshot all'interno della shell HBase	3712
Configurazione di HBase	3713
Modifiche apportate all'allocazione di memoria in YARN	3714
Numeri di porta HBase	3715

Impostazioni sito HBase da ottimizzare	3715
Visualizzazione dell'interfaccia utente di HBase	3717
Visualizzazione di file di log HBase	3719
Monitoraggio di HBase con Ganglia	3720
Migrazione da versioni HBase precedenti	3722
Cronologia dei rilasci di HBase	3722
HCatalog	3788
Creazione di un cluster con HCatalog	3789
Uso di HCatalog	3790
Disabilitazione della scrittura diretta durante l'utilizzo di HCatalog HStorer	3791
Creazione di una tabella tramite la CLI HCat e uso dei dati in Pig	3791
Accesso alla tabella tramite Spark SQL	3792
Esempio: Creare una tabella HCatalog e scrivervi con Pig	3794
Cronologia dei rilasci di HCatalog	3795
Hive	3859
Differenze e considerazioni per Hive su Amazon EMR	3861
Differenze tra Apache Hive su Amazon EMR e Apache Hive	3861
Differenze in Hive tra le versioni Amazon EMR 4.x e 5.x	3862
Caratteristiche aggiuntive di Hive su Amazon EMR	3863
Configurazione di un metastore esterno per Hive	3869
Utilizzo di AWS Glue Data Catalog come metastore Hive	3869
Utilizzo di un database MySQL o Amazon Aurora esterno	3877
Utilizzo del driver JDBC Hive	3879
Miglioramento delle prestazioni Hive	3882
Abilitazione del committer ottimizzato EMRFS S3 di Hive	3882
Uso di S3 Select	3883
Ottimizzazione MSCK	3886
Utilizzo di Hive LLAP	3887
Abilitazione di LLAP	3887
Avvio di LLAP nel cluster	3888
Controllo dello stato di LLAP	3889
Avvio o arresto di LLAP	3889
Ridimensionamento del numero di daemon LLAP	3889
Crittografia in Hive	3890
Crittografia modulare Parquet in Hive	3890
Crittografia in transito in HS2	3894

Cronologia dei rilasci di Hive	3895
Note sul rilascio di Hive per versione	3960
Hudi	4041
Funzionamento di Hudi	4043
Informazioni sui tipi di archiviazione del set di dati: copia in scrittura rispetto a unisci in lettura	4043
Registrazione di un set di dati Hudi con il metastore	4044
Considerazioni e limitazioni	4045
Creazione di un cluster con Hudi installato	4046
Lavorare con un set di dati Hudi	4047
Inizializzazione di una sessione Spark per Hudi	4051
Scrittura in un set di dati Hudi	4051
Upsert dei dati	4057
Eliminazione di un registro	4058
Lettura da un set di dati Hudi	4059
Utilizzo della CLI Hudi	4061
Cronologia dei rilasci di Hudi	4062
Hue	4065
Caratteristiche supportate e non supportate di Hue su Amazon EMR	4067
Connessione all'interfaccia utente Web di Hue	4067
Utilizzo di Hue con un database remoto in Amazon RDS	4068
Risoluzione dei problemi	4070
Configurazioni avanzate per Hue	4071
Configurazione di Hue per utenti LDAP	4071
Cronologia dei rilasci di Hue	4074
Iceberg	4133
Funzionamento di Iceberg	4133
Creazione di un cluster con Iceberg	4135
Utilizza un cluster Iceberg con Spark	4135
Utilizza un cluster Iceberg con Trino	4140
Utilizzo di un cluster Iceberg con Flink	4142
Utilizzo di un cluster Iceberg con Hive	4147
Considerazioni e limitazioni	4150
Considerazioni sull'utilizzo di Iceberg con Spark	4150
Considerazioni sull'utilizzo di Iceberg con Trino	4151
Considerazioni sull'utilizzo di Iceberg con Flink	4151

Considerazioni sull'utilizzo di Iceberg con Hive	4151
Cronologia dei rilasci di Iceberg	4152
Note sul rilascio di Iceberg per versione	4153
Jupyter Notebook	4155
EMR Studio	4155
Notebook EMR	4155
JupyterHub	4155
Creazione di un cluster con JupyterHub	4160
Considerazioni sull'utilizzo di JupyterHub su Amazon EMR	4161
Configurazione di JupyterHub	4162
Configurazione della persistenza per i notebook in Amazon S3	4164
Connessione al nodo master e ai server notebook	4165
Configurazione e amministrazione di JupyterHub	4165
Aggiunta di utenti e amministratori Jupyter Notebook	4167
Installazione di kernel e librerie aggiuntive	4179
Cronologia dei rilasci di JupyterHub	4183
Livy	4217
Abilitazione di HTTPS	4218
Cronologia dei rilasci di Livy	4220
MXNet	4261
Cronologia dei rilasci di MXNet	4262
Oozie	4290
Utilizzo di Oozie con un database remoto in Amazon RDS	4292
Configurazione della versione Java per Oozie	4294
Cronologia dei rilasci di Oozie	4295
Phoenix	4350
Creazione di un cluster con Phoenix	4352
Personalizzazione delle configurazioni Phoenix	4353
Client Phoenix	4354
Cronologia dei rilasci di Phoenix	4358
Pig	4425
Invio di lavoro Pig	4427
Invio di lavoro Pig mediante la console di Amazon EMR	4427
Invio di lavoro Pig mediante la AWS CLI	4428
Chiamata di funzioni UDF (definite dall'utente) da Pig	4430
Chiamata di file JAR da Pig	4430

Chiamata di script Python/Jython da Pig	4430
Cronologia dei rilasci di Pig	4431
Presto e Trino	4492
Utilizzo di Presto con AWS Glue Data Catalog	4494
Specifica di AWS Glue Data Catalog come metastore	4495
Autorizzazioni IAM	3873
Considerazioni sull'utilizzo di AWS Glue Data Catalog	4500
Utilizzo di pushdown S3 Select	4501
Pushdown S3 Select è adatto alla mia applicazione?	4501
Considerazioni e limitazioni	4502
Abilitare Pushdown S3 Select con PrestoDB o Trino	4502
Aggiunta di connettori di database	4503
Utilizzo di SSL/TLS e LDAPS	4504
Utilizzo dell'autenticazione LDAP	4505
Attivazione della modalità rigorosa Presto	4513
Considerazioni	4515
Gestione della perdita di istanze Spot in Presto	4515
Esecuzione con tolleranza ai guasti	4517
Configurazione	4518
Exchange Manager	4518
Considerazioni e limitazioni	4520
Utilizzo della scalabilità automatica Presto con disattivazione con tolleranza	4520
Considerazioni riguardanti Presto on Amazon EMR	4521
Eseguibile della riga di comando Presto	4522
Proprietà di implementazione Presto non configurabili	4522
Installazione di PrestoDB e Trino	4523
Configurazione EMRFS e PrestoS3FileSystem	4523
Impostazione predefinita per rappresentazione utente finale	4524
Porta predefinita per interfaccia Web Presto	4525
Problema con l'esecuzione di bucket Hive in alcuni rilasci	4525
Cronologia delle versioni di Presto	4525
Note di rilascio di Trino (PrestoSQL) per versione	4586
Spark	4589
Creazione di un cluster con Spark	4592
Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.x	4594
Considerazioni sull'esecuzione di Spark con Docker	4595

Creazione di un'immagine Docker	4596
Utilizzo delle immagini Docker da Amazon ECR	4597
Utilizzare AWS Glue Data Catalog come metastore Spark SQL	4602
Specifica di AWS Glue Data Catalog come metastore	4603
Autorizzazioni IAM	3873
Considerazioni sull'utilizzo di AWS Glue Data Catalog	3875
Configurazione di Spark	4608
Valori di default di Spark impostati da Amazon EMR	4609
Configurazione della garbage collection di Spark su Amazon EMR 6.1.0	4611
Utilizzo di maximizeResourceAllocation	4611
Configurazione del comportamento di disattivazione di nodi	4613
Variabile di ambiente ThriftServer Spark	4617
Modifica delle impostazioni di default di Spark	4617
Migrazione da Apache Log4j 1.x a Log4j 2.x	4620
Ottimizzazione delle prestazioni di Spark	4620
Esecuzione adattiva delle query	4621
Eliminazione delle partizioni dinamiche	4623
Appiattimento delle subquery scalari	4625
DISTINCT prima di INTERSECT	4626
Join del filtro Bloom	4627
Riordinamento ottimizzato dei join	4628
Caching dei frammenti dei risultati	4628
Abilitazione del Caching dei frammenti dei risultati Spark	4629
Considerazioni su quando utilizzare il Caching dei frammenti	4630
Utilizzare l'Nvidia Spark-RAPIDS Accelerator per Spark	4632
Scelta dei tipi di istanza	4632
Impostazione delle configurazioni delle applicazioni per il cluster	4632
Aggiunta di un'operazione di bootstrap per il cluster	4637
Avvio del cluster	4637
Accesso alla shell Spark	4638
Utilizzare Amazon SageMaker Spark per il machine learning	4639
Scrittura di un'applicazione Spark	4640
Scala	4640
Java	4641
Python	4642
Migliorare le prestazioni di Spark con Amazon S3	4643

Usare S3 Select	4644
Utilizzare il committer ottimizzato S3 EMRFS	4648
Utilizzo del protocollo di commit ottimizzato per S3 EMRFS	4655
Effettuare nuovi tentativi per le richieste S3	4662
Aggiungi una fase Spark	4665
Sostituzione delle impostazioni di configurazione predefinite di Spark	4668
Visualizzazione della cronologia delle applicazioni Spark	4669
Accedi alle interfacce utente Web di Spark	4669
Utilizzo di Spark su Amazon Redshift	4670
Avvio di un'applicazione Spark	4670
Autenticazione su Amazon Redshift	4671
Lettura e scrittura su Amazon Redshift	4674
Considerazioni	4675
Cronologia dei rilasci di Spark	4676
Sqoop	4736
Considerazioni riguardanti Sqoop su Amazon EMR	4737
Utilizzo di Sqoop con l'integrazione HCatalog	4737
Supporto di Sqoop JDBC e database	4738
Cronologia dei rilasci di Sqoop	4739
TensorFlow	4789
Build di TensorFlow per tipo di istanza Amazon EC2	4790
Sicurezza	4791
Utilizzo di TensorBoard	4791
Cronologia delle versioni di TensorFlow	4792
Tez	4813
Creazione di un cluster con Tez	4814
Configurazione di Tez	4815
Interfaccia utente Web di Tez	4817
Timeline Server	4818
Cronologia dei rilasci di Tez	4818
Note sul rilascio di Tez per versione	4859
Zeppelin	4865
Considerazioni sull'utilizzo di Zeppelin su Amazon EMR	4866
Cronologia dei rilasci di Zeppelin	4867
ZooKeeper	4922
Cronologia dei rilasci di ZooKeeper	4923

Connettori e utility	4960
Esportazione, esecuzione di query e unione di tabelle in DynamoDB	4960
Impostazione di una tabella Hive per l'esecuzione dei comandi Hive	4962
Esempi di comandi Hive per l'esportazione, l'importazione e l'esecuzione di query sui dati	4971
Ottimizzazione delle prestazioni	4980
Kinesis	4984
Cosa posso fare con l'integrazione di Amazon EMR e Amazon Kinesis?	4984
Analisi con checkpoint dei flussi Amazon Kinesis	4985
Considerazioni sulle prestazioni	4986
Pianificazione dell'analisi Amazon Kinesis con Amazon EMR	4986
S3DistCp (s3-dist-cp)	4987
Opzioni S3DistCp	4987
Aggiunta di S3DistCp come fase in un cluster	4995
Eliminazione dopo attività S3DistCp non riuscite	4997
Eseguire comandi e script su un cluster	4999
Invia un passaggio JAR personalizzato per l'esecuzione di uno script o di un comando	4999
Altri modi per utilizzare <code>command-runner.jar</code>	5001
Glossario per AWS	5003

Informazioni sui rilasci di Amazon EMR

Un rilascio di Amazon EMR è un insieme di applicazioni open source dell'ecosistema di Big Data. Ogni rilascio comprende diverse applicazioni, componenti e funzioni di Big Data che si sceglie di avere per installare e configurare Amazon EMR quando si crea un cluster. Le applicazioni vengono raggruppate utilizzando un sistema basato su [Apache BigTop](#), un progetto open-source associato all'ecosistema Hadoop. Questa guida fornisce informazioni per le applicazioni incluse nei rilasci di Amazon EMR.

Per ulteriori informazioni sulle operazioni di base con Amazon EMR, consulta la [Guida per la gestione di Amazon EMR](#).

Quando si avvia un cluster, è possibile scegliere tra diversi rilasci di Amazon EMR. In questo modo l'utente può testare e utilizzare le versioni dell'applicazione che soddisfano i propri requisiti di compatibilità. È possibile specificare la versione di rilascio con l'etichetta del rilascio. Le etichette di rilascio sono sotto forma di `emr-x.x.x`. For example, `emr-6.14.0`.

A partire da Amazon EMR rilascio 5.18.0, è possibile utilizzare i repository degli artefatti di Amazon EMR per creare il codice di processo secondo le versioni esatte di librerie e dipendenze disponibili con rilasci Amazon EMR specifici. Per ulteriori informazioni, consulta [Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR](#).

Iscriviti al feed RSS per le note di rilascio di Amazon EMR all'indirizzo <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/amazon-emr-release-notes.rss> per ricevere aggiornamenti quando è disponibile un nuovo rilascio di Amazon EMR.

Dettagli sul rilascio più recente, incluse le versioni dell'applicazione, le note di rilascio, i componenti e le classificazioni di configurazione di Amazon EMR serie 6.x e 5.x:

- [Rilascio 6.14.0 di Amazon EMR](#)
- [Amazon EMR versione 5.36.1](#)

Note

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Note di rilascio per i rilasci più recenti di Amazon EMR e una cronologia di tutti i rilasci:

- [Novità](#)
- [Archivio delle note di rilascio di Amazon EMR](#)

Una cronologia completa di tutte le versioni dell'applicazione in ogni rilascio di Amazon EMR:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Dettagli per ogni rilascio di Amazon EMR e differenze tra le serie, dove applicabile:

- [Versioni di rilascio di Amazon EMR 6.x](#)
- [Versioni del rilascio di Amazon EMR 5.x](#)
- [Versioni del rilascio di Amazon EMR 4.x](#)
- [Versioni AMI 2.x e 3.x di Amazon EMR](#)

Versioni di rilascio di Amazon EMR 6.x

Questa sezione contiene le versioni dell'applicazione, le note di rilascio, le versioni dei componenti e le classificazioni di configurazione disponibili in ogni versione di rilascio di Amazon EMR 6x.

Quando si avvia un cluster, è possibile scegliere tra diversi rilasci di Amazon EMR. In questo modo l'utente può testare e utilizzare le versioni dell'applicazione che soddisfano i propri requisiti di compatibilità. È possibile specificare la versione di rilascio con l'etichetta del rilascio. Le etichette di rilascio sono sotto forma di `emr-x.x.x`. For example, `emr-6.14.0`.

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Per una tabella completa delle versioni delle applicazioni in ogni versione di Amazon EMR 6.x, consulta [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#).

Argomenti

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Rilascio 6.14.0 di Amazon EMR](#)
- [Versione 6.13.0 di Amazon EMR](#)
- [Versione 6.12.0 di Amazon EMR](#)
- [Versione 6.11.1 di Amazon EMR](#)
- [Versione 6.11.0 di Amazon EMR](#)
- [Versione 6.10.1 di Amazon EMR](#)
- [Versione 6.10.0 di Amazon EMR](#)
- [Versione 6.9.1 di Amazon EMR](#)
- [Amazon EMR rilascio 6.9.0](#)
- [Versione 6.8.1 di Amazon EMR](#)
- [Amazon EMR versione 6.8.0](#)
- [Rilascio 6.7.0 di Amazon EMR](#)
- [Rilascio 6.6.0 di Amazon EMR](#)
- [Rilascio 6.5.0 di Amazon EMR](#)
- [Rilascio 6.4.0 di Amazon EMR](#)
- [Rilascio 6.3.1 di Amazon EMR](#)
- [Rilascio 6.3.0 di Amazon EMR](#)
- [Rilascio 6.2.1 di Amazon EMR](#)
- [Rilascio 6.2.0 di Amazon EMR](#)
- [Rilascio 6.1.1 di Amazon EMR](#)
- [Rilascio 6.1.0 di Amazon EMR](#)
- [Rilascio 6.0.1 di Amazon EMR](#)
- [Rilascio 6.0.0 di Amazon EMR](#)

Versioni dell'applicazione nei rilasci di Amazon EMR 6.x

Per una tabella completa che elenca le versioni dell'applicazione disponibili in ogni versione di Amazon EMR 6.x, aprire [Versioni delle applicazioni nelle versioni di Amazon EMR 6.x](#) nel browser.

Rilascio 6.14.0 di Amazon EMR

- [Versioni delle applicazioni](#)

- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
SDK AWS per 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
Delta	2.4.0	2.4.0	2.4.0	2.2.0
Flink	1,17.1-amzn-0	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

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-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
Presto	0.281	0.281	0.281	0.279
Spark	3.4.1	3.4.1	3.4.0	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.14.0 di Amazon EMR. Le modifiche sono relative alla versione 6.13.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.
- La [scalabilità gestita di Amazon EMR](#) è ora disponibile nella regione ap-southeast-3 Asia Pacific (Giacarta) per i cluster creati con Amazon EMR 6.14.0 e versioni successive.

Modifiche, miglioramenti e problemi risolti

- La versione 6.14.0 ottimizza la gestione dei log con Amazon EMR in esecuzione su Amazon EC2. Di conseguenza, potresti notare una leggera riduzione dei costi di archiviazione per i log del cluster.
- La versione 6.14.0 migliora il flusso di lavoro di dimensionamento per tenere conto delle diverse istanze principali che presentano una variazione sostanziale delle dimensioni dei loro volumi Amazon EBS. Questo miglioramento si applica solo ai nodi principali; le operazioni di riduzione delle dimensioni per i nodi di attività non sono influenzate.
- La versione 6.14.0 migliora il modo in cui Amazon EMR interagisce con applicazioni open source come Apache Hadoop YARN ResourceManager and HDFS NameNode. Questo miglioramento riduce il rischio di ritardi operativi dovuti nel dimensionamento del cluster e mitiga gli errori di avvio dovuti a problemi di connettività con le applicazioni open source.
- La versione 6.14.0 ottimizza l'installazione delle applicazioni all'avvio del cluster. Ciò migliora i tempi di avvio del cluster per determinate combinazioni di applicazioni Amazon EMR.

- La versione 6.14.0 corregge un problema a causa del quale le operazioni di riduzione del cluster potrebbero bloccarsi quando un cluster in esecuzione in un VPC con un dominio personalizzato rileva un riavvio del nodo principale o del nodo attività.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023.906.0	4.14.322	11 settembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			e), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker

Componente	Versione	Descrizione
delta	2.4.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
delta-standalone-connectors	0.6.0	I connettori Delta offrono runtime diversi per integrare Delta Lake con motori come Flink, Hive e Presto.
emr-ddb	5.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.7.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.11.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.28.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.7.0	Connettore di EMR S3Select
emr-wal-cli	1.1.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.59.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1,17.1-amzn-0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1,17.1-amzn-0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-journalnode	3.3.3-amzn-6	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.3.3-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.17-amzn-2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	2.4.17-amzn-2	Servizio per servire una o più regioni HBase.
hbase-client	2.4.17-amzn-2	Client di riga di comando HBase.
hbase-rest-server	2.4.17-amzn-2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.17-amzn-2	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.17-amzn-2	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-7	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-7	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-7	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-7	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-7	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	3.1.3-amzn-7	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-7	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.13.1-amzn-2	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.13.1-amzn-2	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.13.1-amzn-2	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.13.1-amzn-2	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.3.1-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.5.0	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Le librerie Phoenix per server e client
phoenix-connectors	5.1.3	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.281-amzn-2	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.281-amzn-2	Servizio per l'esecuzione di parti di una query.
presto-client	0.281-amzn-2	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	422-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	422-amzn-0	Servizio per l'esecuzione di parti di una query.
trino-client	422-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.4.1-amzn-1	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	3.4.1-amzn-1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.4.1-amzn-1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.4.1-amzn-1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	23.06.0-amzn-2	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-5	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-5	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.14.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifica la home Java KMS di Hadoop	Not available.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.

Classificazioni	Descrizione	Operazioni di riconfigurazione
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-lakeformat ion	Modifica i valori nel file lakeformation.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
trino-log	Modifica i valori nel file log.properties di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-kafka	Modifica i valori nel file kafka.properties di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file memory.properties di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file mongodb.properties di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file mysql.properties di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file postgresql.properties di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file raptor.properties di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file redis.properties di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file redshift.properties di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file kms-logback.xml di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.14.0 e note di rilascio

Data	Evento	Descrizione
02/11/2023	Implementazione completata	Amazon EMR 6.14.0 implementato completamente in tutte le regioni supportate
10/10/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.14.0
04/10/2023	Versione iniziale	Amazon EMR 6.14.0 è stato implementato per la prima volta in regioni commerciali limitate

Versione 6.13.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

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

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.281	0.279	0.279
Spark	3.4.1	3.4.0	3.3.2	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.13.0 di Amazon EMR.

Le modifiche sono relative alla versione 6.12.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.

Modifiche, miglioramenti e problemi risolti

- La versione 6.13.0 migliora il daemon di gestione dei log di Amazon EMR per garantire che tutti i log vengano caricati a cadenza regolare su Amazon S3 quando viene emesso un comando di terminazione del cluster. Ciò facilita le terminazioni più rapide dei cluster.
- La versione 6.13.0 migliora le funzionalità di gestione dei log di Amazon EMR per garantire il caricamento coerente e tempestivo di tutti i file di log su Amazon S3. Ciò avvantaggia in particolare i cluster EMR di lunga durata.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-808.0	4.14.320	24 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.2	SDK Spark Amazon SageMaker
delta	2.4.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
delta-standalone-connectors	0.6.0	I connettori Delta offrono runtime diversi per integrare Delta Lake con motori come Flink, Hive e Presto.
emr-ddb	5.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.6.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.10.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.27.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.6.0	Connettore di EMR S3Select
emr-wal-cli	1.1.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.

Componente	Versione	Descrizione
emrfs	2.58.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.17.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.17.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-5	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	3.3.3-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-5	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.3.3-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	2.4.17-amzn-1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.17-amzn-1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.17-amzn-1	Client di riga di comando HBase.
hbase-rest-server	2.4.17-amzn-1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.17-amzn-1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.17-amzn-1	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-6	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-6	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-6	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-6	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-hbase	3.1.3-amzn-6	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-6	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-6	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.13.1-amzn-1	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.13.1-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.13.1-amzn-1	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.13.1-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.3.0-amzn-1	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.5.0	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Le librerie Phoenix per server e client
phoenix-connectors	5.1.3	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.281-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.281-amzn-1	Servizio per l'esecuzione di parti di una query.
presto-client	0.281-amzn-1	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	414-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	414-amzn-1	Servizio per l'esecuzione di parti di una query.
trino-client	414-amzn-1	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	3.4.1-amzn-0	Client a riga di comando Spark.
spark-history-server	3.4.1-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.4.1-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.4.1-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	23.06.0-amzn-1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-4	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-4	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.13.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifica la home Java KMS di Hadoop	Not available.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.13.0 e note di rilascio

Data	Evento	Descrizione
23/09/2023	Implementazione completata	Amazon EMR 6.13.0 implementato completamente in tutte le regioni supportate
12/09/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.13.0
01/09/2023	Versione iniziale	Implementazione di Amazon EMR 6.13.0 in alcune regioni commerciali

Versione 6.12.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

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

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.10.0
Iceberg	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.279	0.279	0.278
Spark	3.4.0	3.3.2	3.3.2	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	410	410	403

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.12.0 di Amazon EMR. Le modifiche sono relative alla versione 6.11.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.
- Le versioni 6.12.0 e successive di Amazon EMR supportano l'integrazione LDAP con Apache Livy, Apache Hive tramite HiveServer2 (HS2), Trino, Presto e Hue. Inoltre, puoi installare Apache Spark e Apache Hadoop su un cluster EMR che utilizza 6.12.0 o versioni successive e configurarli per l'utilizzo di LDAP. Per ulteriori informazioni, consulta [Utilizzo dei server Active Directory o LDAP per l'autenticazione con Amazon EMR](#).

Modifiche, miglioramenti e problemi risolti

- Le versioni 6.12.0 e successive di Amazon EMR forniscono il supporto di runtime Java 11 per Flink. Per ulteriori informazioni, consulta [Configurazione di Flink per l'esecuzione con Java 11](#).
- Amazon EMR 6.12.0 supporta tutte le applicazioni con Amazon Corretto 8 per impostazione predefinita, ad eccezione di Trino. Per Trino, Amazon EMR supporta Amazon Corretto 17 per impostazione predefinita a partire dalla versione 6.9.0 di Amazon EMR. Amazon EMR supporta anche alcune applicazioni con Amazon Corretto 11 e 17. Queste applicazioni sono elencate nella seguente tabella. Se desideri modificare la JVM predefinita sul cluster, segui le istruzioni riportate in [Configurazione delle applicazioni per l'uso di una macchina virtuale Java specifica](#) per ogni applicazione in esecuzione sul cluster. Puoi utilizzare solo una delle versioni di runtime Java per cluster. Amazon EMR non supporta l'esecuzione di nodi o applicazioni differenti su versioni di runtime diverse sullo stesso cluster.

Sebbene Amazon EMR supporti Amazon Corretto 11 e 17 su Apache Spark, Apache Hadoop e Apache Hive, le prestazioni potrebbero peggiorare per alcuni carichi di lavoro quando utilizzi queste versioni di Corretto. Consigliamo di testare i carichi di lavoro prima di modificare i valori predefiniti.

Versioni Java predefinite per le applicazioni in Amazon EMR 6.12

Applicazione	Versione Java/Amazon Corretto (l'impostazione predefinita è in grassetto)
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
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8

Applicazione	Versione Java/Amazon Corretto (l'impostazione predefinita è in grassetto)
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

- La versione 6.12.0 aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.
- La versione 6.12.0 corregge un problema a causa del quale le operazioni di riduzione del cluster potrebbero bloccarsi quando un nodo principale sottoposto a una disattivazione graduale perde integrità per qualsiasi motivo prima della completa disattivazione.
- La versione 6.12.0 migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- La versione 6.12.0 migliora le prestazioni e l'efficienza del servizio di monitoraggio dell'integrità per Amazon EMR aumentando la velocità con cui registra le modifiche di stato per le istanze. Questo miglioramento riduce la possibilità di un peggioramento delle prestazioni per i nodi del cluster che eseguono più strumenti client personalizzati o applicazioni di terze parti.
- La versione 6.12.0 migliora le prestazioni del daemon di gestione dei log sul cluster per Amazon EMR. Di conseguenza, ci sono meno possibilità di ridurre le prestazioni con i cluster EMR che eseguono fasi con elevata simultaneità.
- Con la versione 6.12.0 di Amazon EMR, il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.

- La versione 6.12.0 include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- La versione 6.12.0 consente la rotazione dei log di YARN Timeline Server. Ciò riduce al minimo gli scenari di utilizzo eccessivo del disco, in particolare per i cluster di lunga durata.
- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.4.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>delta-standalone-connectors</code>	0.6.0	I connettori Delta offrono runtime diversi per integrare Delta Lake con motori come Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.5.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.9.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.26.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.5.0	Connettore di EMR S3Select
emr-wal-cli	1.1.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.57.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.17.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.17.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.3.3-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.3.3-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.17-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.17-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.4.17-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.4.17-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.17-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.17-amzn-0	Strumento di riparazione per cluster Apache HBase.

Componente	Versione	Descrizione
hcatalog-client	3.1.3-amzn-5	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-5	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-5	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-5	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-5	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-5	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-5	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.13.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.13.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.13.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.

Componente	Versione	Descrizione
hudi-spark	0.13.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.3.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Le librerie Phoenix per server e client
phoenix-connectors	5.1.3	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.281-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.281-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.281-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	414-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	414-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	414-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.4.0-amzn-0	Client a riga di comando Spark.
spark-history-server	3.4.0-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.4.0-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.4.0-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	23.06.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-3	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-3	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.12.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifica la home Java KMS di Hadoop	Not available.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.12.0 e note di rilascio

Data	Evento	Descrizione
27/07/2023	Aggiorna documentazione	Aggiorna le opzioni di Java per la versione 6.12 e aggiungi il tutorial di Oozie per aggiornar e la JVM
21/07/2023	Implementazione completata	Amazon EMR 6.12.0 implementato completamente in tutte le regioni supportate
21/07/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.12.0
12/07/2023	Versione iniziale	Implementazione di Amazon EMR 6.12.0 in alcune regioni commerciali

Versione 6.11.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
SDK AWS per Java	1.12.446	1.12.446	1.12.397	1.12.397
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
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

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
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
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

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.11.1 di Amazon EMR. Le modifiche sono relative alla versione 6.11.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- A causa di un conflitto di blocco, un nodo può entrare in una situazione di stallo se viene aggiunto o rimosso contemporaneamente al tentativo di disattivazione. Di conseguenza, Hadoop Resource Manager (YARN) non risponde e influisce su tutti i container in entrata e quelli attualmente in esecuzione.
- Questa versione include una modifica che consente ai cluster ad alta disponibilità di ripristinare lo stato di errore dopo il riavvio.
- Questa versione include correzioni di sicurezza per Hue e HBase.
- Questa versione corregge un problema per cui i cluster che eseguono carichi di lavoro su Spark con Amazon EMR potrebbero ricevere silenziosamente risultati errati con `contains`, `startsWith`, `endsWith` e `like`. Questo problema si verifica quando utilizzi le espressioni su campi partizionati che contengono metadati in Amazon EMR Hive3 Metastore Server (HMS).
- Questa versione corregge un problema di limitazione (della larghezza di banda della rete) sul lato Glue quando non sono presenti funzioni definite dall'utente (UDF).
- Questa versione corregge un problema che comporta l'eliminazione dei log dei container tramite il servizio di aggregazione dei log dei nodi prima che il pusher possa inviarli a S3 in caso di disattivazione di YARN.
- Questa versione corregge un problema con i parametri di FairShare Scheduler quando Node Label è abilitato per Hadoop.

- Questa versione corregge un problema che influiva sulle prestazioni di Spark quando veniva impostato un valore `true` predefinito per la configurazione di `spark.yarn.heterogeneousExecutors.enabled` in `spark-defaults.conf`.
- Questa versione corregge un problema a causa del quale Reduce Task non era in grado di leggere i dati shuffle. Il problema ha causato errori nelle query di Hive con un errore di memoria danneggiata.
- Questa versione aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.
- Questa versione migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- Il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- Questa versione include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.2.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>delta-standalone-connectors</code>	0.6.0	I connettori Delta offrono runtime diversi per integrare Delta Lake con motori come Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.8.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.25.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.4.0	Connettore di EMR S3Select
emr-wal-cli	1.1.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.56.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.16.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.16.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-3.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-3.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-3.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-3.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.3.3-amzn-3.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-3.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-3.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-3.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.15-amzn-1.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.15-amzn-1.1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.15-amzn-1.1	Client di riga di comando HBase.
hbase-rest-server	2.4.15-amzn-1.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.15-amzn-1.1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.15-amzn-1.1	Strumento di riparazione per cluster Apache HBase.

Componente	Versione	Descrizione
hcatalog-client	3.1.3-amzn-4.1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-4.1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-4.1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-4.1	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-4.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-4.1	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-4.1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.13.0-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.13.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.13.0-amzn-0	Libreria bundle per eseguire Trino con Hudi.

Componente	Versione	Descrizione
hudi-spark	0.13.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.279-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.279-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.279-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	410-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti trino-workers.
trino-worker	410-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	410-amzn-0	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.2-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.3.2-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.2-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.2-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	23.02.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-2.1	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-2.1	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.11.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.11.1 e note di rilascio

Data	Evento	Descrizione
30/08/2023	Aggiorna le note di rilascio	Aggiunte diverse correzioni relative al piano di controllo alle note di rilascio
21/08/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.11.1
16/08/2023	Implementazione completata	Amazon EMR 6.11.1 implementato completamente in tutte le regioni supportate
04/08/2023	Versione iniziale	Implementazione di Amazon EMR 6.11.1 in alcune regioni commerciali

Versione 6.11.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

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

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

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.11.0 di Amazon EMR. Le modifiche sono relative alla versione 6.10.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove funzionalità di

- Amazon EMR 6.11.0 supporta 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 e PrestoDB 0.279-amzn-0.

Modifiche, miglioramenti e problemi risolti

- Con Amazon EMR 6.11.0, il connettore DynamoDB è stato aggiornato alla versione 5.0.0. La versione 5.0.0 utilizza AWS SDK for Java 2.x. Le versioni precedenti utilizzavano AWS SDK for Java 1.x. Come risultato di questo aggiornamento, consigliamo vivamente di testare il codice prima di utilizzare il connettore DynamoDB con Amazon EMR 6.11.
- Quando il connettore DynamoDB per Amazon EMR 6.11.0 chiama il servizio DynamoDB, utilizza il valore della regione fornito per la proprietà di `dynamodb.endpoint`. Consigliamo di configurare anche `dynamodb.region` quando utilizzi `dynamodb.endpoint` e che entrambe le proprietà abbiano come destinazione la stessa Regione AWS. Se utilizzi `dynamodb.endpoint` e non configuri `dynamodb.region`, il connettore DynamoDB per Amazon EMR 6.11.0 restituisce un'eccezione della regione non valida e tenta di riconciliare le informazioni della Regione AWS dal servizio di metadati di istanza (IMDS) Amazon EC2. Se il connettore non è in grado di recuperare la regione da IMDS, il valore predefinito è Stati Uniti orientali (Virginia settentrionale) (`us-east-1`). L'errore seguente è un esempio dell'eccezione della regione non valida che

potresti ottenere se non configuri correttamente la proprietà `dynamodb.region`: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Per ulteriori informazioni sulle classi interessate dall'aggiornamento di AWS SDK for Java a 2.x, consulta il commit [Aggiornamento di AWS SDK for Java da 1.x a 2.x \(#175\)](#) nel repository GitHub per il connettore Amazon EMR - DynamoDB.

- Questa versione corregge un problema per cui i dati diventano NULL quando utilizzi Delta Lake per archiviare i dati delle tabelle Delta in Amazon S3 dopo l'operazione di ridenominazione delle colonne. Per ulteriori informazioni su questa funzionalità sperimentale di Delta Lake, consulta [Operazione di ridenominazione delle colonne](#) nella Guida per l'utente di Delta Lake.
- La versione 6.11.0 corregge un problema che potrebbe verificarsi quando si crea un nodo edge replicando uno dei nodi primari da un cluster con più nodi primari. Il nodo edge replicato potrebbe causare ritardi nelle operazioni di riduzione o comportare un elevato utilizzo della memoria sui nodi primari. Per ulteriori informazioni su come creare un nodo edge per comunicare con il cluster EMR, consulta [Strumento per creare nodi edge](#) nel repository `aws-samples` su GitHub.
- La versione 6.11.0 migliora il processo di automazione utilizzato da Amazon EMR per rimontare i volumi Amazon EBS su un'istanza dopo un riavvio.
- La versione 6.11.0 corregge un problema che causava discontinuità nei parametri Hadoop pubblicati da Amazon EMR su Amazon CloudWatch.
- La versione 6.11.0 corregge un problema con i cluster EMR per cui un aggiornamento del file di configurazione YARN che contiene l'elenco di esclusione dei nodi per il cluster viene interrotto a causa dell'eccessivo utilizzo del disco. L'aggiornamento incompleto ostacola le future operazioni di riduzione del cluster. Questa versione garantisce l'integrità del cluster e il corretto funzionamento delle operazioni di dimensionamento.
- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con le versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` è impostato su `false` in `yarn-site.xml` per risolvere questo problema.

Sebbene la correzione si occupi dei problemi introdotti da YARN-9608, potrebbe influire sulla corretta esecuzione dei processi di Hive a causa della perdita dei dati shuffle sui cluster con Scalabilità gestita abilitata. In questa versione abbiamo mitigato tale rischio impostando anche `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` per i carichi di lavoro di Hive. Questa configurazione è disponibile solo con Amazon EMR 6.11.0 e versioni successive.

- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-0727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.2.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>delta-standalone-connectors</code>	0.6.0	I connettori Delta offrono runtime diversi per integrare Delta Lake con motori come Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.8.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.25.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.4.0	Connettore di EMR S3Select
emr-wal-cli	1.1.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.56.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.16.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.16.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-3	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.3.3-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.3.3-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.15-amzn-1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.15-amzn-1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.15-amzn-1	Client di riga di comando HBase.
hbase-rest-server	2.4.15-amzn-1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.15-amzn-1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.15-amzn-1	Strumento di riparazione per cluster Apache HBase.

Componente	Versione	Descrizione
hcatalog-client	3.1.3-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-4	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-4	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.13.0-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.13.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.13.0-amzn-0	Libreria bundle per eseguire Trino con Hudi.

Componente	Versione	Descrizione
hudi-spark	0.13.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.279-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.279-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.279-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	410-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti trino-workers.
trino-worker	410-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	410-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.2-amzn-0	Client a riga di comando Spark.
spark-history-server	3.3.2-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.2-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.2-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	23.02.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-2	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-2	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.11.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the Resource Manager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.11.0 e note di rilascio

Data	Evento	Descrizione
21/08/2023	Aggiornamento	Correzione del problema introdotto con Hadoop 3.3.3.
26/07/2023	Aggiornamento	Nuove etichette della versione del sistema operativo 2.0.20230612.0 e 2.0.20230628.0 .
09/06/2023	Implementazione completata	Amazon EMR 6.11.0 implementato completamente in tutte le regioni supportate
09/06/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.11.0

Data	Evento	Descrizione
08/06/2023	Versione iniziale	Implementazione di Amazon EMR 6.11.0 in alcune regioni commerciali

Versione 6.10.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
SDK AWS per Java	1.12.397	1.12.397	1.12.170	1.12.170

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.10.1 di Amazon EMR. Le modifiche sono relative alla versione 6.10.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- A causa di un conflitto di blocco, un nodo può entrare in una situazione di stallo se viene aggiunto o rimosso contemporaneamente al tentativo di disattivazione. Di conseguenza, Hadoop Resource Manager (YARN) non risponde e influisce su tutti i container in entrata e quelli attualmente in esecuzione.
- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con le versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` è impostato su `false` in `yarn-site.xml` per risolvere questo problema.

Sebbene la correzione si occupi dei problemi introdotti da YARN-9608, potrebbe influire sulla corretta esecuzione dei processi di Hive a causa della perdita dei dati shuffle sui cluster con Scalabilità gestita abilitata. In questa versione abbiamo mitigato tale rischio impostando anche `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` per i carichi di lavoro di Hive. Questa configurazione è disponibile solo con Amazon EMR 6.11.0 e versioni successive.

- Metrics Collector non invia alcun parametro al piano di controllo dopo il failover del nodo primario nei cluster con la configurazione dei gruppi di istanze.
- Questa versione include una modifica che consente ai cluster ad alta disponibilità di ripristinare lo stato di errore dopo il riavvio.
- Questa versione include correzioni di sicurezza per Hue e HBase.
- Questa versione corregge un problema per cui i cluster che eseguono carichi di lavoro su Spark con Amazon EMR potrebbero ricevere silenziosamente risultati errati con `contains`, `startsWith`, `endsWith` e `like`. Questo problema si verifica quando utilizzi le espressioni su campi partizionati che contengono metadati in Amazon EMR Hive3 Metastore Server (HMS).
- Questa versione corregge un problema di limitazione (della larghezza di banda della rete) sul lato Glue quando non sono presenti funzioni definite dall'utente (UDF).
- Questa versione corregge un problema che comporta l'eliminazione dei log dei container tramite il servizio di aggregazione dei log dei nodi prima che il pusher possa inviarli a S3 in caso di disattivazione di YARN.
- Questa versione corregge un problema con i parametri di FairShare Scheduler quando Node Label è abilitato per Hadoop.
- Questa versione corregge un problema che influiva sulle prestazioni di Spark quando veniva impostato un valore `true` predefinito per la configurazione di `spark.yarn.heterogeneousExecutors.enabled` in `spark-defaults.conf`.
- Questa versione corregge un problema a causa del quale Reduce Task non era in grado di leggere i dati shuffle. Il problema ha causato errori nelle query di Hive con un errore di memoria danneggiata.
- Questa versione aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il

ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.

- Questa versione migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- Il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- Questa versione include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- Questa versione corregge un problema che potrebbe verificarsi quando si crea un nodo edge replicando uno dei nodi primari da un cluster con più nodi primari. Il nodo edge replicato potrebbe causare ritardi nelle operazioni di riduzione o comportare un elevato utilizzo della memoria sui nodi primari. Per ulteriori informazioni su come creare un nodo edge per comunicare con il cluster EMR, consulta [Strumento per creare nodi edge](#) nel repository `aws-samples` su GitHub.
- Questa versione migliora il processo di automazione utilizzato da Amazon EMR per rimontare i volumi Amazon EBS su un'istanza dopo un riavvio.
- Questa versione corregge un problema che causava discontinuità nei parametri Hadoop pubblicati da Amazon EMR su Amazon CloudWatch.
- Questa versione corregge un problema con i cluster EMR per cui un aggiornamento del file di configurazione YARN che contiene l'elenco di esclusione dei nodi per il cluster viene interrotto a causa dell'eccessivo utilizzo del disco. L'aggiornamento incompleto ostacola le future operazioni di riduzione del cluster. Questa versione garantisce l'integrità del cluster e il corretto funzionamento delle operazioni di dimensionamento.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-0727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.2.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.7.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter

Componente	Versione	Descrizione
emr-s3-dist-cp	2.24.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.3.0	Connettore di EMR S3Select
emr-wal-cli	1.0.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.55.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.16.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.16.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.3.3-amzn-2.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-2.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-2.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-2.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.3.3-amzn-2.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-2.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-2.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.15-amzn-0.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.15-amzn-0.1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.15-amzn-0.1	Client di riga di comando HBase.
hbase-rest-server	2.4.15-amzn-0.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.15-amzn-0.1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.15-amzn-0.1	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-3.1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.3-amzn-3.1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-3.1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-3.1	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-3.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-3.1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-3.1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.12.2-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.12.2-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.12.2-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.12.2-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.1.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.5.0	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	6.0.0-SNAPSHOT	Connettori Apache Phoenix per Spark-3
phoenix-query-server	6.0.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.278.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.278.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.278.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	403-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	403-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	403-amzn-0	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.1-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.3.1-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.1-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.1-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.12.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-1.1	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-1.1	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.10.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oracoop-site	Modifica i valori nel file oracoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.10.1 e note di rilascio

Data	Evento	Descrizione
30/08/2023	Aggiorna le note di rilascio	Aggiunte diverse correzioni relative al piano di controllo alle note di rilascio
21/08/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.10.1
16/08/2023	Implementazione completata	Amazon EMR 6.10.1 implementato completamente in tutte le regioni supportate
04/08/2023	Versione iniziale	Implementazione di Amazon EMR 6.10.1 in alcune regioni commerciali

Versione 6.10.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
SDK AWS per 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
Delta	2.2.0	2.1.0	2.1.0	-
Flink	1.16.0	1.15.2	1.15.2	1.15.1

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

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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.10.0 di Amazon EMR. Le modifiche sono relative alla versione 6.9.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove funzionalità di

- Amazon EMR 6.10.0 supporta Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 e PrestoDB 0.278.1.
- Amazon EMR 6.10.0 include un connettore Trino-Hudi nativo che fornisce l'accesso in lettura ai dati nelle tabelle Hudi. Puoi attivare il connettore con `trino-cli --catalog hudi` e configurarlo in base alle tue esigenze con `trino-connector-hudi`. L'integrazione nativa con Amazon EMR indica che non è più necessario utilizzare `trino-connector-hive` per eseguire query sulle tabelle Hudi. Per un elenco delle configurazioni supportate con il nuovo connettore, consulta la pagina [Connettori Hudi](#) nella documentazione di Trino.
- Le versioni 6.10.0 e successive di Amazon EMR supportano l'integrazione di Apache Zeppelin con Apache Flink. Per ulteriori informazioni, consulta [Utilizzo dei processi Flink di Zeppelin in Amazon EMR](#).

Problemi noti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario

eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Per risolvere questo problema in Amazon EMR 6.10.0, puoi impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Amazon EMR 6.10.0 rimuove la dipendenza da `minimal-json.jar` per l'[integrazione Amazon Redshift per Apache Spark](#) e aggiunge automaticamente i jar necessari relativi a Spark-Redshift al percorso della classe dell'esecutore per Spark: `spark-redshift.jar`, `spark-avro.jar` e `RedshiftJDBC.jar`.
- La versione 6.10.0 migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- La versione 6.10.0 riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- Amazon EMR 6.10.0 supporta gli endpoint regionali per la mappatura degli utenti EMRFS.
- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- La versione 6.10.0 corregge un problema che causava il blocco dei processi Spark quando tutti gli esecutori Spark rimanenti si trovavano su un host in fase di disattivazione con il gestore di risorse YARN.
- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.

- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.2.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.7.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter

Componente	Versione	Descrizione
emr-s3-dist-cp	2.24.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.3.0	Connettore di EMR S3Select
emr-wal-cli	1.0.0	Interfaccia della linea di comando utilizzata per l'elenco o la cancellazione di emrwal.
emrfs	2.55.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.16.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.16.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.3.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-2	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.3.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.3.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.15-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.15-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.4.15-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.4.15-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.15-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.15-amzn-0	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.3-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-3	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-3	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.12.2-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.12.2-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.12.2-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.12.2-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	1.1.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.5.0	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.8.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	6.0.0-SNAPSHOT	Connettori Apache Phoenix per Spark-3
phoenix-query-server	6.0.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.278.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.278.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.278.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	403-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti trino-workers.
trino-worker	403-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	403-amzn-0	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.1-amzn-0	Client a riga di comando Spark.
spark-history-server	3.3.1-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.1-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.1-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.12.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-1	L'applicazione e le librerie tez YARN.
tez-on-worker	0.10.2-amzn-1	L'applicazione e le librerie tez YARN per i nodi worker.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.10.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the Resource Manager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-lakeformation	Modifica i valori nel file <code>lakeformation.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file <code>delta.properties</code> di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file <code>exchange-manager.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifica i valori nel file <code>hudi.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifica i valori nel file <code>kms-logback.xml</code> di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.10.0 e note di rilascio

Data	Evento	Descrizione
21/08/2023	Aggiornamento	Aggiunto un problema noto introdotto con Hadoop 3.3.3.
26/07/2023	Aggiornamento	Nuove etichette della versione del sistema operativo 2.0.20230612.0 e 2.0.20230628.0 .
02/03/2023	Implementazione completata	Amazon EMR 6.10 implementato completamente in tutte le regioni supportate
02/03/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.10

Data	Evento	Descrizione
27/02/2023	Versione iniziale	Implementazione di Amazon EMR 6.10 in alcune regioni commerciali

Versione 6.9.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
SDK AWS per Java	1.12.170	1.12.170	1.12.170	1.12.170

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.9.1 di Amazon EMR. Le modifiche sono relative alla versione 6.9.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con le versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` è impostato su `false` in `yarn-site.xml` per risolvere questo problema.

Sebbene la correzione si occupi dei problemi introdotti da YARN-9608, potrebbe influire sulla corretta esecuzione dei processi di Hive a causa della perdita dei dati shuffle sui cluster con Scalabilità gestita abilitata. In questa versione abbiamo mitigato tale rischio impostando anche

`yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` per i carichi di lavoro di Hive. Questa configurazione è disponibile solo con Amazon EMR 6.11.0 e versioni successive.

- Metrics Collector non invia alcun parametro al piano di controllo dopo il failover del nodo primario nei cluster con la configurazione dei gruppi di istanze.
- Questa versione elimina i nuovi tentativi sulle richieste HTTP con esito negativo agli endpoint di Metrics Collector.
- Questa versione include una modifica che consente ai cluster ad alta disponibilità di ripristinare lo stato di errore dopo il riavvio.
- Questa versione corregge un problema per cui gli UID di grandi dimensioni creati dall'utente causavano eccezioni di overflow.
- Questa versione corregge i problemi di timeout con il processo di riconfigurazione di Amazon EMR.
- Questa versione include correzioni di sicurezza.
- Questa versione corregge un problema per cui i cluster che eseguono carichi di lavoro su Spark con Amazon EMR potrebbero ricevere silenziosamente risultati errati con `contains`, `startsWith`, `endsWith` e `like`. Questo problema si verifica quando utilizzi le espressioni su campi partizionati che contengono metadati in Amazon EMR Hive3 Metastore Server (HMS).
- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.
- Hive potrebbe subire una perdita di dati quando utilizzi HDFS come directory scratch, hai abilitato l'unione di file di piccole dimensioni e la tabella contiene percorsi di partizione statici.
- Questa versione corregge un problema di prestazioni con Hive se l'unione di file di piccole dimensioni (disabilitata per impostazione predefinita) è abilitata alla fine del processo ETL.
- Questa versione corregge un problema di limitazione (della larghezza di banda della rete) sul lato Glue quando non sono presenti funzioni definite dall'utente (UDF).
- Questa versione corregge un problema che comporta l'eliminazione dei log dei container tramite il servizio di aggregazione dei log dei nodi prima che il pusher possa inviarli a S3 in caso di disattivazione di YARN.

- Questa versione corregge la gestione dei file compattati/archiviati con il monitoraggio persistente degli storefile per HBase.
- Questa versione corregge un problema che influiva sulle prestazioni di Spark quando veniva impostato un valore `true` predefinito per la configurazione di `spark.yarn.heterogeneousExecutors.enabled` in `spark-defaults.conf`.
- Questa versione corregge un problema a causa del quale Reduce Task non era in grado di leggere i dati shuffle. Il problema ha causato errori nelle query di Hive con un errore di memoria danneggiata.
- Questa versione corregge un problema che causava l'esito negativo del provider del nodo se il servizio HDFS NameNode (NN) era bloccato in safemode durante la sostituzione del nodo.
- Questa versione aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.
- Questa versione migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- Il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- Questa versione include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- Questa versione corregge un problema che potrebbe verificarsi quando si crea un nodo edge replicando uno dei nodi primari da un cluster con più nodi primari. Il nodo edge replicato potrebbe causare ritardi nelle operazioni di riduzione o comportare un elevato utilizzo della memoria sui nodi primari. Per ulteriori informazioni su come creare un nodo edge per comunicare con il cluster EMR, consulta [Strumento per creare nodi edge](#) nel repository `aws-samples` su GitHub.
- Questa versione migliora il processo di automazione utilizzato da Amazon EMR per rimontare i volumi Amazon EBS su un'istanza dopo un riavvio.

- Questa versione corregge un problema che causava discontinuità nei parametri Hadoop pubblicati da Amazon EMR su Amazon CloudWatch.
- Questa versione corregge un problema con i cluster EMR per cui un aggiornamento del file di configurazione YARN che contiene l'elenco di esclusione dei nodi per il cluster viene interrotto a causa dell'eccessivo utilizzo del disco. L'aggiornamento incompleto ostacola le future operazioni di riduzione del cluster. Questa versione garantisce l'integrità del cluster e il corretto funzionamento delle operazioni di dimensionamento.
- Questa versione migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- Questa versione riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.2	SDK Spark Amazon SageMaker
delta	2.1.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
emr-ddb	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.6.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.23.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.2.0	Connettore di EMR S3Select
emrfs	2.54.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.15.2	Applicazioni e script client a riga di comando Apache Flink.

Componente	Versione	Descrizione
flink-jobmanager-config	1.15.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-0.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-0.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.3.3-amzn-0.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-0.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-journalnode	3.3.3-amzn-0.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.3.3-amzn-0.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-0.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-0.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-0.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-0.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-0.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.13-amzn-0.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	2.4.13-amzn-0.1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.13-amzn-0.1	Client di riga di comando HBase.
hbase-rest-server	2.4.13-amzn-0.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.13-amzn-0.1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.13-amzn-0.1	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-2.1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-2.1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-2.1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-2.1	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-2.1	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	3.1.3-amzn-2.1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-2.1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.12.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.12.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.12.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.12.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.7.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	6.0.0-SNAPSHOT	Connettori Apache Phoenix per Spark-3
phoenix-query-server	6.0.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.276-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.276-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.276-amzn-0	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	398-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti trino-workers.
trino-worker	398-amzn-0	Servizio per l'esecuzione di parti di una query.
trino-client	398-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.0-amzn-1.1	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	3.3.0-amzn-1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.0-amzn-1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.0-amzn-1.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.08.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.10.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-0.1	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.9.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.
presto-connector-lakeformat ion	Modifica i valori nel file lakeformation.properties di Presto.	Restarts Presto-Server (for PrestoDB)

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
trino-log	Modifica i valori nel file log.properties di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file config.properties di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-password-authenticator	Modifica i valori nel file password-authenticator.properties di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file trino-env.sh di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file node.properties di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file blackhole.properties di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file cassandra.properties di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file delta.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifica i valori nel file hive.properties di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file exchange-manager.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file iceberg.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file jmx.properties di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file kafka.properties di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.9.1 e note di rilascio

Data	Evento	Descrizione
30/08/2023	Aggiorna le note di rilascio	Aggiunte diverse correzioni relative al piano di controllo alle note di rilascio
21/08/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.9.1
16/08/2023	Implementazione completata	Amazon EMR 6.9.1 implementato completamente in tutte le regioni supportate
04/08/2023	Versione iniziale	Implementazione di Amazon EMR 6.9.1 in alcune regioni commerciali

Amazon EMR rilascio 6.9.0

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
SDK AWS per 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
Delta	2.1.0	-	-	-
Flink	1.15.2	1.15.1	1.15.1	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.12	2.4.12	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.276	0.273	0.273	0.272
Spark	3.3.0	3.3.0	3.3.0	3.2.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.9.1	2.9.1	2.4.1
Tez	0.10.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.9.0. Le modifiche sono relative al rilascio di Amazon EMR 6.8.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove caratteristiche

- Amazon EMR rilascio 6.9.0 supporta Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 e Tez 0.10.2.
- Amazon EMR rilascio 6.9.0 include una nuova applicazione open source, [Delta Lake](#) 2.1.0.
- L'integrazione di Amazon Redshift per Apache Spark è inclusa in Amazon EMR rilascio 6.9.0 e successivi. In precedenza uno strumento open source, l'integrazione nativa è un connettore Spark che è possibile utilizzare per creare applicazioni Apache Spark in grado di leggere e scrivere dati in Amazon Redshift e Amazon Redshift Serverless. Per ulteriori informazioni, consulta [Utilizzo dell'integrazione di Amazon Redshift per Apache Spark su Amazon EMR](#).
- Amazon EMR rilascio 6.9.0 aggiunge il supporto per l'archiviazione dei log su Amazon S3 durante la riduzione dei cluster. In precedenza, era possibile archiviare i file di log in Amazon S3 solo durante la terminazione del cluster. La nuova funzionalità garantisce la persistenza su Amazon S3 dei file di log generati nel cluster anche dopo la terminazione del nodo. Per ulteriori informazioni, consulta [Configurazione della registrazione e del debug di cluster](#).
- Per supportare le query di lunga durata, Trino ora include un meccanismo di esecuzione con tolleranza ai guasti. L'esecuzione con tolleranza ai guasti mitiga gli errori delle query ritentando le query non riuscite o le attività dei loro componenti. Per ulteriori informazioni, consulta [Esecuzione con tolleranza ai guasti in Trino](#).
- È possibile utilizzare Apache Flink su Amazon EMR per l'elaborazione BATCH e STREAM unificata delle tabelle Apache Hive o dei metadati di qualsiasi origine di tabella Flink, come Iceberg, Kinesis o Kafka. È possibile specificare AWS Glue Data Catalog come metastore per Flink utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Per ulteriori informazioni, consulta [Configurazione di Flink in Amazon EMR](#).
- Ora è possibile specificare i ruoli di runtime AWS Identity and Access Management (IAM) e il controllo degli accessi basato su AWS Lake Formation per Apache Spark, Apache Hive e Presto su Amazon EMR sui cluster EC2 con Amazon SageMaker Studio. Per ulteriori informazioni, consulta [Configurazione dei ruoli di runtime per le fasi di Amazon EMR](#).

Problemi noti

- Per Amazon EMR rilascio 6.9.0, Trino non funziona su cluster abilitati per Apache Ranger. Se hai la necessità di utilizzare Trino con Ranger, contatta il [AWS Support](#).
- Se utilizzi l'integrazione di Amazon Redshift per Apache Spark e disponi di un'indicazione temporale time, timez, timestamp o timestampz con una precisione di microsecondi in formato

Parquet, il connettore arrotonda i valori temporali al valore in millisecondi più vicino. Come soluzione alternativa, utilizza il parametro `unload_s3_format` del formato di scaricamento del testo.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Le connessioni ai cluster Amazon EMR da Amazon SageMaker Studio possono fallire in modo intermittente con un codice di risposta 403 Forbidden. Questo errore si verifica quando l'impostazione del ruolo IAM sul cluster richiede più di 60 secondi. Come soluzione alternativa, puoi installare una patch Amazon EMR per consentire nuovi tentativi e aumentare il timeout a un minimo di 300 secondi. Completa la seguente procedura per applicare l'operazione bootstrap all'avvio del cluster.

1. Scarica lo script di bootstrap e i file RPM tramite i seguenti URL Amazon S3.

```
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. Carica i file della fase precedente in un bucket Amazon S3 di tua proprietà. Il bucket deve trovarsi nella stessa Regione AWS in cui intendi avviare il cluster.

- Includi la seguente operazione di bootstrap durante l'avvio del cluster EMR. Sostituisci *bootstrap_URI* e *RPM_URI* con gli URI corrispondenti di Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi SecretAgent e RecordServer servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà Log4j2. Ca sua della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink fornisce connettori nativi per i file system S3 e Hadoop che consentono alle applicazioni di creare un FileSink e di scrivere i dati in Amazon S3. Questo FileSink fallisce con una delle due eccezioni seguenti.

```
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:?]
```

Come soluzione alternativa, puoi installare una patch Amazon EMR, che risolve il problema sopra riportato in Flink. Per applicare l'operazione bootstrap all'avvio del cluster, completa la seguente procedura.

1. Scarica il [flink-rpm](#) nel bucket Amazon S3. Il tuo percorso RPM è `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Scarica lo script di bootstrap e i file RPM da Amazon S3 utilizzando il seguente URI. Sostituisci *regionName* con la Regione AWS in cui prevedi di avviare il cluster.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-rpms.sh
```

3. Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. In Amazon EMR 6.8.0 e 6.9.0, questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con [Amazon EMR 6.10.0](#), esiste una soluzione alternativa a questo problema per impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Per la release 6.9.0 e successive di Amazon EMR, tutti i componenti installati da Amazon EMR che utilizzano librerie Log4j utilizzano Log4j versione 2.17.1 o successiva.
- Quando utilizzi il connettore DynamoDB con Spark nelle versioni 6.6.0, 6.7.0 e 6.8.0 di Amazon EMR, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Amazon EMR rilascio 6.9.0 risolve questo problema.
- Amazon EMR 6.9.0 aggiunge un supporto limitato per il controllo degli accessi basato su Lake Formation con Apache Hudi durante la lettura dei dati tramite Spark SQL. Il supporto è per le query SELECT che utilizzano Spark SQL ed è limitato al controllo degli accessi a livello di colonna. Per ulteriori informazioni, consulta [Hudi e Formation](#).

- Quando usi Amazon EMR 6.9.0 per creare un cluster Hadoop con le [Etichette nodo](#) abilitate, l'[API delle metriche YARN](#) restituisce informazioni aggregate su tutte le partizioni, anziché sulla partizione predefinita. Per ulteriori informazioni, consulta [YARN-11414](#).
- Con Amazon EMR 6.9.0, abbiamo aggiornato Trino alla versione 398, che utilizza Java 17. La versione precedente supportata di Trino per Amazon EMR 6.8.0 era Trino 388 in esecuzione su Java 11. Per ulteriori informazioni su questa modifica, consulta [Aggiornamenti di Trino a Java 17](#) sul blog di Trino.
- Questa versione corregge un problema di mancata corrispondenza della sequenza temporale tra Apache BigTop e Amazon EMR nella sequenza di avvio del cluster EC2. Questa mancata corrispondenza della sequenza temporale si verifica quando un sistema tenta di eseguire due o più operazioni contemporaneamente anziché eseguirle nella sequenza corretta. Di conseguenza, alcune configurazioni del cluster hanno registrato timeout di avvio delle istanze e tempi di avvio del cluster più lenti.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	(Centrale), Israele (Tel Aviv) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	12 gennaio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR

e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>delta</code>	2.1.0	Delta Lake è un formato a tabella aperta per set di dati analitici di grandissime dimensioni
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.6.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
<code>emr-s3-dist-cp</code>	2.23.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emr-s3-select	2.2.0	Connettore di EMR S3Select
emrfs	2.54.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.15.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.15.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.3.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.3.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	3.3.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.3.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.3.3-amzn-1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	3.3.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.3.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.3.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.3.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.3.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.3.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	2.4.13-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.13-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.4.13-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.4.13-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.13-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.13-amzn-0	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-2	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-hbase	3.1.3-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.12.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.12.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.12.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.12.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.7.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	6.0.0-SNAPSHOT	Connettori Apache Phoenix per Spark-3
phoenix-query-server	6.0.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.276-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.276-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.276-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	398-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	398-amzn-0	Servizio per l'esecuzione di parti di una query.
trino-client	398-amzn-0	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	3.3.0-amzn-1	Client a riga di comando Spark.
spark-history-server	3.3.0-amzn-1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.0-amzn-1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.0-amzn-1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.08.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.10.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.10.2-amzn-0	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.9.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
delta-defaults	Modifica i valori nel file delta-defaults.conf di Delta.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j2	Modifica le impostazioni di log4j2.properties di Livy.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.
presto-connector-lakeformat ion	Modifica i valori nel file lakeformation.properties di Presto.	Restarts Presto-Server (for PrestoDB)

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
trino-log	Modifica i valori nel file log.properties di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file config.properties di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-password-authenticator	Modifica i valori nel file password-authenticator.properties di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file trino-env.sh di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file node.properties di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file blackhole.properties di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file cassandra.properties di Trino.	Not available.
trino-connector-delta	Modifica i valori nel file delta.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifica i valori nel file hive.properties di Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modifica i valori nel file exchange-manager.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file iceberg.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file jmx.properties di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file kafka.properties di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per il rilascio 6.9.0 e le note di rilascio

Data	Evento	Descrizione
30/08/2023	Aggiorna le note di rilascio	Aggiunta una correzione per il problema di mancata corrispondenza della sequenza temporale
21/08/2023	Aggiorna le note di rilascio	Aggiunto un problema noto con Hadoop 3.3.3.
26/07/2023	Aggiornamento	Nuove etichette della versione del sistema operativo 2.0.20230612.0 e 2.0.20230628.0 .
13/12/2022	Aggiornamento delle note di rilascio	Aggiunta di una caratteristica e di un problema di runtime noto con SageMaker
29/11/2022	Aggiornamento delle note di rilascio e della documentazione	Aggiunta di una caratteristica per l'integrazione di Amazon Redshift per Apache Spark
23/11/2022	Aggiornamento delle note di rilascio	Rimozione della voce Log4j
18/11/2022	Implementazione completata	Amazon EMR 6.9 implementato completamente in tutte le regioni supportate
18/11/2022	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.9
14/11/2022	Versione iniziale	Implementazione di Amazon EMR 6.9 in alcune regioni commerciali

Versione 6.8.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
SDK AWS per 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
Delta	-	-	-	-
Flink	1.15.1	1.15.1	1.14.2	1.14.2

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.12	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.2
Hudi	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.273	0.273	0.272	0.267
Spark	3.3.0	3.3.0	3.2.1	3.2.0
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

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 6.8.1 di Amazon EMR. Le modifiche sono relative alla versione 6.8.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con le versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` è impostato su `false` in `yarn-site.xml` per risolvere questo problema.

Sebbene la correzione si occupi dei problemi introdotti da YARN-9608, potrebbe influire sulla corretta esecuzione dei processi di Hive a causa della perdita dei dati shuffle sui cluster con Scalabilità gestita abilitata. In questa versione abbiamo mitigato tale rischio impostando anche `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` per i carichi di lavoro di Hive. Questa configurazione è disponibile solo con Amazon EMR 6.11.0 e versioni successive.

- Metrics Collector non invia alcun parametro al piano di controllo dopo il failover del nodo primario nei cluster con la configurazione dei gruppi di istanze.

- Questa versione elimina i nuovi tentativi sulle richieste HTTP con esito negativo agli endpoint di Metrics Collector.
- Questa versione include una modifica che consente ai cluster ad alta disponibilità di ripristinare lo stato di errore dopo il riavvio.
- Questa versione corregge un problema per cui gli UID di grandi dimensioni creati dall'utente causavano eccezioni di overflow.
- Questa versione corregge i problemi di timeout con il processo di riconfigurazione di Amazon EMR.
- Questa versione previene un problema per cui una riconfigurazione con esito negativo potrebbe interrompere altri processi non correlati.
- Questa versione include correzioni di sicurezza.
- Questa versione corregge un problema per cui i cluster che eseguono carichi di lavoro su Spark con Amazon EMR potrebbero ricevere silenziosamente risultati errati con `contains`, `startsWith`, `endsWith` e `like`. Questo problema si verifica quando utilizzi le espressioni su campi partizionati che contengono metadati in Amazon EMR Hive3 Metastore Server (HMS).
- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.
- Hive potrebbe subire una perdita di dati quando utilizzi HDFS come directory scratch, hai abilitato l'unione di file di piccole dimensioni e la tabella contiene percorsi di partizione statici.
- Questa versione corregge un problema di prestazioni con Hive se l'unione di file di piccole dimensioni (disabilitata per impostazione predefinita) è abilitata alla fine del processo ETL.
- Questa versione corregge un problema di limitazione (della larghezza di banda della rete) sul lato Glue quando non sono presenti funzioni definite dall'utente (UDF).
- Questa versione corregge un problema che comporta l'eliminazione dei log dei container tramite il servizio di aggregazione dei log dei nodi prima che il pusher possa inviarli a S3 in caso di disattivazione di YARN.
- Questa versione corregge la gestione dei file compattati/archiviati con il monitoraggio persistente degli storefile per HBase.

- Questa versione corregge un problema che influiva sulle prestazioni di Spark quando veniva impostato un valore `true` predefinito per la configurazione di `spark.yarn.heterogeneousExecutors.enabled` in `spark-defaults.conf`.
- Questa versione corregge un problema a causa del quale Reduce Task non era in grado di leggere i dati shuffle. Il problema ha causato errori nelle query di Hive con un errore di memoria danneggiata.
- Questa versione corregge un problema che causava l'esito negativo del provider del nodo se il servizio HDFS NameNode (NN) era bloccato in `safemode` durante la sostituzione del nodo.
- Questa versione aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.
- Questa versione migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- Il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- Questa versione include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- Questa versione corregge un problema che potrebbe verificarsi quando si crea un nodo edge replicando uno dei nodi primari da un cluster con più nodi primari. Il nodo edge replicato potrebbe causare ritardi nelle operazioni di riduzione o comportare un elevato utilizzo della memoria sui nodi primari. Per ulteriori informazioni su come creare un nodo edge per comunicare con il cluster EMR, consulta [Strumento per creare nodi edge](#) nel repository `aws-samples` su GitHub.
- Questa versione migliora il processo di automazione utilizzato da Amazon EMR per rimontare i volumi Amazon EBS su un'istanza dopo un riavvio.
- Questa versione corregge un problema che causava discontinuità nei parametri Hadoop pubblicati da Amazon EMR su Amazon CloudWatch.

- Questa versione corregge un problema con i cluster EMR per cui un aggiornamento del file di configurazione YARN che contiene l'elenco di esclusione dei nodi per il cluster viene interrotto a causa dell'eccessivo utilizzo del disco. L'aggiornamento incompleto ostacola le future operazioni di riduzione del cluster. Questa versione garantisce l'integrità del cluster e il corretto funzionamento delle operazioni di dimensionamento.
- Questa versione migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- Questa versione riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- Questa versione aggiunge il supporto per l'archiviazione dei log su Amazon S3 durante la riduzione dei cluster. In precedenza, era possibile archiviare i file di log in Amazon S3 solo durante la terminazione del cluster. La nuova funzionalità garantisce la persistenza su Amazon S3 dei file di log generati nel cluster anche dopo la terminazione del nodo. Per ulteriori informazioni, consulta [Configurazione della registrazione e del debug di cluster](#).
- Questa versione corregge un problema che si verificava quando l'URI di Amazon S3 per un'operazione di bootstrap terminava con un numero di porta, ad esempio: a . b . c . d : 4345. Amazon EMR stava analizzando erroneamente questi URI, quindi tutte le operazioni di bootstrap associate avrebbero avuto esito negativo.
- Questa versione corregge un problema di mancata corrispondenza della sequenza temporale tra Apache BigTop e Amazon EMR nella sequenza di avvio del cluster EC2. Questa mancata corrispondenza della sequenza temporale si verifica quando un sistema tenta di eseguire due o più operazioni contemporaneamente anziché eseguirle nella sequenza corretta. Di conseguenza, alcune configurazioni del cluster hanno registrato timeout di avvio delle istanze e tempi di avvio del cluster più lenti.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
<code>emr-s3-dist-cp</code>	2.22.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	2.1.0	Connettore di EMR S3Select

Componente	Versione	Descrizione
emrfs	2.53.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.15.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.15.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-8.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-8.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-8.1	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	3.2.1-amzn-8.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.2.1-amzn-8.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-8.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-8.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	2.4.12-amzn-0.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.12-amzn-0.1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.12-amzn-0.1	Client di riga di comando HBase.
hbase-rest-server	2.4.12-amzn-0.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.12-amzn-0.1	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.12-amzn-0.1	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-1.1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-1.1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-1.1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-1.1	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-hbase	3.1.3-amzn-1.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-1.1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-1.1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.11.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.11.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.11.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.11.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.14.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.7.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.273.3-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.273.3-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.273.3-amzn-0	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	388-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	388-amzn-0	Servizio per l'esecuzione di parti di una query.
trino-client	388-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	3.3.0-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.3.0-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.0-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.0-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.06.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.9.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.8.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the Resource Manager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
trino-log	Modifica i valori nel file log.properties di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file config.properties di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file password-authenticator.properties di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file trino-env.sh di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file node.properties di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file blackhole.properties di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file <code>sqoop-site.xml</code> di Sqoop.	Not available.
tez-site	Modifica i valori nel file <code>tez-site.xml</code> file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file <code>yarn-site.xml</code> di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in <code>zeppelin-site.xml</code> .	Restarts Zeppelin.

Classificazioni	Descrizione	Operazioni di riconfigurazione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.8.1 e note di rilascio

Data	Evento	Descrizione
30/08/2023	Aggiorna le note di rilascio	Aggiunte diverse correzioni relative al piano di controllo alle note di rilascio
21/08/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 6.8.1
16/08/2023	Implementazione completata	Amazon EMR 6.8.1 implementato completamente in tutte le regioni supportate
04/08/2023	Versione iniziale	Implementazione di Amazon EMR 6.8.1 in alcune regioni commerciali

Amazon EMR versione 6.8.0

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Liviy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
SDK AWS per 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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.8.0. Le modifiche sono relative alla versione 6.7.0

Nuove caratteristiche

- Le fasi di Amazon EMR supportano gli endpoint Apache Livy e i client JDBC/ODBC. Per ulteriori informazioni, consulta [Configurazione dei ruoli di runtime per le fasi di Amazon EMR](#).
- Il rilascio 6.8.0 di Amazon EMR include Spark per Apache HBase 2.4.12. Con questa versione di HBase, è possibile archiviare ed eliminare le tabelle HBase. Il processo di archiviazione di Amazon S3 rinomina tutti i file della tabella nella directory di archivio. Questo processo può essere lungo e costoso. Ora è possibile saltare il processo di archiviazione ed eliminare rapidamente tabelle di grandi dimensioni. Per ulteriori informazioni, consulta [Utilizzo della shell HBase](#).

Problemi noti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. In Amazon EMR 6.8.0 e 6.9.0, questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con [Amazon EMR 6.10.0](#), esiste una soluzione alternativa a questo problema per impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Quando i rilasci 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggevano le tabelle di Apache Phoenix tramite la shell Apache Spark, Amazon EMR produceva un `NoSuchMethodError`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- Il rilascio 6.8.0 di Amazon EMR include [Apache Hudi](#) 0.11.1; tuttavia, i cluster Amazon EMR 6.8.0 sono compatibili anche con l'open source `hudi-spark3.3-bundle_2.12` di Hudi 0.12.0.

- Il rilascio 6.8.0 di Amazon EMR con Apache Spark 3.3.0. Questa versione di Spark utilizza Apache Log4j 2 e il file `log4j2.properties` per configurare Log4j nei processi Spark. Se utilizzi Spark nel cluster o crei cluster EMR con parametri di configurazione personalizzati e desideri eseguire l'aggiornamento alla versione 6.8.0 di Amazon EMR, devi migrare alla nuova classificazione della configurazione e formato chiave `spark-log4j2` per Apache Log4j 2. Per ulteriori informazioni, consulta [Migrazione da Apache Log4j 1.x a Log4j 2.x](#).
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	Arabi Uniti), Canada (Centrale) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	6 settembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Problemi noti

- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la

divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. Ca sua della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.2	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-notebook-env	1.7.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.22.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.53.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.15.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.15.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.2.1-amzn-8	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-8	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-8	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-8	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-8	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-8	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-8	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-8	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-8	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.2.1-amzn-8	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-8	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.12-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.12-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.4.12-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.4.12-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.12-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.12-amzn-0	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-1	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.11.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.11.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.11.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.11.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.14.0-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.9.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.7.0	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.273.3-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.273.3-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.273.3-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	388-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	388-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	388-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.3.0-amzn-0	Client a riga di comando Spark.
spark-history-server	3.3.0-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.3.0-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.3.0-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.06.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.9.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.8.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.	Not available.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file memory.properties di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file mongodb.properties di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file mysql.properties di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file postgresql.properties di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file raptor.properties di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file redis.properties di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file redshift.properties di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j2	Modifica i valori nel file log4j2.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 6.8.0 e le note di rilascio

Data	Evento	Descrizione
21/08/2023	Aggiornamento	Aggiunto un problema noto con Hadoop 3.3.3.
26/07/2023	Aggiornamento	Nuove etichette della versione del sistema operativo 2.0.20230612.0 e 2.0.20230628.0 .
06/09/2022	Implementazione completata	Amazon EMR 6.8 implementato completamente in tutte le regioni supportate
06/09/2022	Pubblicazione iniziale	Pubblicazione delle prime note di rilascio di Amazon EMR 6.8
31/08/2022	Versione iniziale	Amazon EMR 6.8 pubblicato in regioni commerciali limitate

Rilascio 6.7.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
SDK AWS per 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
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.0	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.272	0.267	0.261	0.254.1
Spark	3.2.1	3.2.0	3.1.2	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	378	367	360	359
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.5.7

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.7.0. Le modifiche sono relative alla versione 6.6.0.

Data del rilascio iniziale: 15 luglio 2022

Nuove caratteristiche

- Amazon EMR ora supporta Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 e Trino 0.378.
- Supporta i controlli degli accessi basati su ruoli IAM e Lake Formation con fasi EMR (Spark, Hive) per Amazon EMR sui cluster EC2.
- Supporta le istruzioni di definizione dei dati Apache Spark su cluster abilitati Apache Ranger. Adesso, include il supporto per le applicazioni Trino che leggono e scrivono metadati Apache Hive su cluster abilitati Apache Ranger. Per ulteriori informazioni, consulta la sezione [Abilitazione della governance federata utilizzando Trino e Apache Ranger su Amazon EMR](#).
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 719.0	4.14.287	10 agosto 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022.606.1	4.14.281	15 luglio 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché

Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.

- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. A causa della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.2.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.6.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.22.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.52.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.14.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-7	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-7	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-7	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-7	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-7	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.2.1-amzn-7	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-7	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-7	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.2.1-amzn-7	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-7	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.4-amzn-3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.4-amzn-3	Servizio per servire una o più regioni HBase.
hbase-client	2.4.4-amzn-3	Client di riga di comando HBase.
hbase-rest-server	2.4.4-amzn-3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.4-amzn-3	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.4-amzn-3	Strumento di riparazione per cluster Apache HBase.

Componente	Versione	Descrizione
hcatalog-client	3.1.3-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.3-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.3-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.3-amzn-0	Client a riga di comando Hive.
hive-hbase	3.1.3-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-0	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.3-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.11.0-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.11.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.11.0-amzn-0	Libreria bundle per eseguire Trino con Hudi.

Componente	Versione	Descrizione
hudi-spark	0.11.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.13.1-amzn-0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.0.194	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.272-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.272-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.272-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	378-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti trino-workers.
trino-worker	378-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	378-amzn-0	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.2.1-amzn-0	Client a riga di comando Spark.
spark-history-server	3.2.1-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.2.1-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.2.1-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.02.0-amzn-1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.7	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.7	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.7.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.	Not available.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file memory.properties di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file mongodb.properties di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file mysql.properties di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file postgresql.properties di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file raptor.properties di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file redis.properties di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file redshift.properties di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.6.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
SDK AWS per 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
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

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	0.13.1	0.12.0	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.261	0.254.1	0.245.1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
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

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.6.0. Le modifiche sono relative alla versione 6.5.0.

Data del rilascio iniziale: 9 maggio 2022

Ultimo aggiornamento della documentazione: 15 giugno 2022

Nuove caratteristiche

- Amazon EMR 6.6 ora supporta Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 e PrestoDB 0.267.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2022 805.0	4.14.287	30 agosto 2022	us-west-1

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022719.0	4.14.287	10 agosto 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 426.0	4.14.281	10 giugno 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022406.1	4.14.275	2 maggio 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

- Le applicazioni Amazon EMR versione 6.6 e successive che usano Log4j 1.x e Log4j 2.x sono aggiornate per usare rispettivamente Log4j 1.2.17 (o superiore) e Log4j 2.17.1 (o superiore) e non richiedono l'uso di [operazioni di bootstrap](#) per mitigare i problemi CVE.
- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento

gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).

- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.

Modifiche, miglioramenti e problemi risolti

- Amazon EMR riduce in media i tempi di avvio del cluster fino a 80 secondi per i cluster che utilizzano l'opzione AMI di default EMR e installano solo applicazioni comuni, come Apache Hadoop, Apache Spark e Apache Hive.

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.
- Su cluster Trino a lunga esecuzione Amazon EMR 6.6.0 abilita i parametri di registrazione della rimozione di oggetti inutili (Garbage Collection) in Trino `jvm.config` per ottenere informazioni migliori dai log di Garbage Collection. Questa modifica aggiunge molti registri di Garbage Collection al file `launcher.log` (`/var/log/trino/launcher.log`). Se esegui cluster Trino in Amazon EMR 6.6.0, potresti riscontrare che i nodi esauriscono lo spazio su disco dopo che il cluster è stato in esecuzione per un paio di giorni a causa dei registri che sono stati aggiunti.

La soluzione alternativa per questo problema è eseguire lo script sottostante come operazione Bootstrap per disabilitare i parametri di registrazione di rimozione di oggetti inutili (garbage collection) in `jvm.config` durante la creazione o la clonazione del cluster per Amazon EMR 6.6.0.

```
#!/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"
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. A causa della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```


Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-notebook-env	1.5.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.20.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.50.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.14.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.2.1-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-6	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.2.1-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.4-amzn-2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.4-amzn-2	Servizio per servire una o più regioni HBase.
hbase-client	2.4.4-amzn-2	Client di riga di comando HBase.
hbase-rest-server	2.4.4-amzn-2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.4-amzn-2	Servizio che fornisce un endpoint Thrift a HBase.
hbase-operator-tools	2.4.4-amzn-2	Strumento di riparazione per cluster Apache HBase.
hcatalog-client	3.1.2-amzn-7	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.2-amzn-7	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-7	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-7	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-7	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-7	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-7	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.10.1-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.10.1-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.10.1-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.10.1-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
iceberg	0.13.1	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	11.0.194	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	5.1.2	Le librerie Phoenix per server e client
phoenix-connectors	5.1.2	Connettori Apache Phoenix per Spark-3
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.267-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.267-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.267-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	367-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	367-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
trino-client	367-amzn-0	Trino client della riga di comando che viene installato o su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.2.0-amzn-0	Client a riga di comando Spark.
spark-history-server	3.2.0-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.2.0-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.2.0-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	22.02.0-amzn-0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.7	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.7	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.6.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.	Not available.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifica i valori nel file <code>iceberg.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file memory.properties di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file mongodb.properties di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file mysql.properties di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file postgresql.properties di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file raptor.properties di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file redis.properties di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file redshift.properties di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.5.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
SDK AWS per 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
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

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.9.0
Iceberg	0.12.0	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.261	0.254.1	0.245.1	0.245.1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
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

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.5.0. Le modifiche sono relative alla versione 6.4.0.

Data del rilascio iniziale: 20 gennaio 2022

Data di rilascio aggiornata: 21 marzo 2022

Nuove caratteristiche

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).
- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.
- Supporto per il formato di tabella aperta Apache Iceberg per enormi set di dati analitici.
- Supporto per ranger-trino-plugin 2.0.1-amzn-1
- Supporto per toree 0.5.0

Modifiche, miglioramenti e problemi risolti

- La versione del rilascio 6.5 di Amazon EMR ora supporta Apache Iceberg 0.12.0 e offre miglioramenti al tempo di esecuzione con Amazon EMR Runtime per Apache Spark, Amazon EMR Runtime per Presto e Amazon EMR Runtime per Apache Hive.
- [Apache Iceberg](#) è un formato di tabella aperta per set di dati di grandi dimensioni in Amazon S3 e fornisce prestazioni di query rapide su tabelle di grandi dimensioni, commit atomici, scritture

simultanee ed evoluzione delle tabelle compatibili con SQL. Con EMR 6.5, è possibile utilizzare Apache Spark 3.1.2 con il formato tabella Iceberg.

- Apache Hudi 0.9 aggiunge il supporto Spark SQL DDL e DML. Ciò consente di creare e modificare le tabelle Hudi usando solo istruzioni SQL. Apache Hudi 0.9 include anche miglioramenti delle prestazioni sul lato query e sul lato scrittore.
- Amazon EMR Runtime per Apache Hive migliora le prestazioni di Apache Hive su Amazon S3 rimuovendo le operazioni di ridenominazione durante le operazioni di gestione temporanea e migliora le prestazioni per i comandi di metastore check (MSCK) utilizzati per la riparazione delle tabelle.

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- I cluster bundle Hbase in High Availability (HA) non riescono a effettuare il provisioning con la dimensione del volume e il tipo di istanza di default. La soluzione alternativa per questo problema consiste nell'aumentare la dimensione del volume principale.
- Per utilizzare le operazioni Spark con Apache Oozie, devi aggiungere la seguente configurazione al tuo file `workflow.xml` Oozie. In caso contrario, diverse librerie critiche come Hadoop e EMRFS non saranno presenti nella classpath degli esecutori Spark lanciati da Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra a e b in `s3://bucket/table/p=a b`. Tieni presente che esistono

altri 14 caratteri non di controllo: !"#%&'()*+,-. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-notebook-env	1.4.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.19.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.48.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.14.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.2.1-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-5	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.2.1-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.4-amzn-1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.4-amzn-1	Servizio per servire una o più regioni HBase.
hbase-client	2.4.4-amzn-1	Client di riga di comando HBase.
hbase-rest-server	2.4.4-amzn-1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.4-amzn-1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-6	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-6	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	3.1.2-amzn-6	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-6	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-6	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-6	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-6	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.9.0-amzn-1	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-presto	0.9.0-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.9.0-amzn-1	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.9.0-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
iceberg	0.12.0	Apache Iceberg è un formato a tabella aperta per enormi set di dati analitici
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.261-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.261-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.261-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	360	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	360	Servizio per l'esecuzione di parti di una query.
trino-client	360	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing

Componente	Versione	Descrizione
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.1.2-amzn-1	Client a riga di comando Spark.
spark-history-server	3.1.2-amzn-1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.1.2-amzn-1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.1.2-amzn-1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.4.1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.7	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.7	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.5.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
iceberg-defaults	Modifica i valori nel file iceberg-defaults.conf di Iceberg.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.	Not available.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
trino-log	Modifica i valori nel file <code>log.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file <code>config.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file <code>trino-env.sh</code> di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file <code>node.properties</code> di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Trino.	Not available.
trino-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.4.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
SDK AWS per 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
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.2.6	2.2.6	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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
Trino (PrestoSQL)	359	350	350	343

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.4.0. Le modifiche sono relative alla versione 6.3.0.

Data del rilascio iniziale: 20 settembre 2021

Data di rilascio aggiornata: 21 marzo 2022

Applicazioni supportate

- AWS SDK for Java versione 1.12.31
- CloudWatch Sink versione 2.2.0
- DynamoDB Connector versione 4.16.0
- EMRFS versione 2.47.0
- Amazon EMR Goodies versione 3.2.0
- Amazon EMR Kinesis Connector versione 3.5.0
- Amazon EMR Record Server versione 2.1.0
- Amazon EMR Scripts versione 2.5.0
- Flink versione 1.13.1
- Ganglia versione 3.7.2
- AWS Glue Hive Metastore Client versione 3.3.0
- Hadoop versione 3.2.1-amzn-4
- HBase versione 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalog versione 3.1.2-amzn-5
- Hive versione 3.1.2-amzn-5
- Hudi versione 0.8.0-amzn-0

- Hue versione 4.9.0
- Java JDK versione Corretto-8.302.08.1 (build 1.8.0_302-b08)
- JupyterHub versione 1.4.1
- Livy versione 0.7.1-incubating
- MXNet versione 1.8.0
- Oozie versione 5.2.1
- Phoenix versione 5.1.2
- Pig versione 0.17.0
- Presto versione 0.254.1-amzn-0
- Trino versione 359
- Apache Ranger KMS (crittografia trasparente multi-master) versione 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK versione 1.4.1
- Scala versione 2.12.10 (VM server OpenJDK a 64 bit, Java 1.8.0_282)
- Spark versione 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop versione 1.4.7
- TensorFlow versione 2.4.1
- tez versione 0.9.2
- Zeppelin versione 0.9.0
- Zookeeper versione 3.5.7
- Connettori e driver: DynamoDB Connector 4.16.0

Nuove funzionalità di

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).

- Sui cluster Amazon EMR abilitati ad Apache Ranger, puoi utilizzare Apache Spark SQL per inserire o aggiornare i dati nelle tabelle dei metastore di Apache Hive utilizzando INSERT INTO, INSERT OVERWRITE, e ALTER TABLE. Quando si utilizza ALTER TABLE con Spark SQL, una posizione di partizione deve essere la directory figlio di una posizione di tabella. Al momento Amazon EMR non supporta l'inserimento di dati in una partizione in cui la posizione della partizione è diversa da quella della tabella.
- PrestoSQL è stato [rinominato in Trino](#).
- Hive: l'esecuzione di semplici query SELECT con clausola LIMIT viene accelerata interrompendo l'esecuzione della query non appena viene recuperato il numero di registri menzionati nella clausola LIMIT. Le query SELECT semplici sono query che non hanno una clausola GROUP BY/ORDER by o query che non hanno uno stadio di riduzione. Ad esempio, SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Controlli di simultaneità Hudi

- Hudi ora supporta Optimistic Concurrency Control (OCC), che può essere sfruttato con operazioni di scrittura come UPSERT e INSERT per consentire modifiche da più scrittori alla stessa tabella Hudi. Questo è livello di file OCC, quindi due commit (o scrittori) possono scrivere sulla stessa tabella, se le modifiche non sono in conflitto. Per ulteriori informazioni, consulta [Controllo della concorrenza Hudi](#).
- I cluster Amazon EMR hanno installato Zookeeper, che può essere sfruttato come provider di blocchi per OCC. Per semplificare l'utilizzo di questa funzione, i cluster Amazon EMR hanno le seguenti proprietà preconfigurate:

```
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
```

Per abilitare OCC, è necessario configurare le seguenti proprietà con le relative opzioni di processo Hudi o a livello di cluster utilizzando l'API di configurazione Amazon EMR:

```
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)
```

Monitoraggio Hudi: integrazione di Amazon CloudWatch per segnalare i parametri Hudi

- Amazon EMR supporta la pubblicazione di parametri Hudi su Amazon CloudWatch. È abilitato impostando le seguenti configurazioni richieste:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Di seguito sono riportate le configurazioni Hudi opzionali che è possibile modificare:

Impostazione	Descrizione	Valore
hoodie.metrics.cloudwatch.report.period.seconds	Frequenza (in secondi) con cui segnalare i parametri ad Amazon CloudWatch	Il valore di default è 60s, il che va bene per la risoluzione di default di un minuto offerta da Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefisso da aggiungere a ciascun nome parametro	Il valore di default è vuoto (nessun prefisso)
hoodie.metrics.cloudwatch.namespace	Spazio dei nomi Amazon CloudWatch in base al quale vengono pubblicati i parametri	Il valore di default è Hudi
hoodie.metrics.cloudwatch.maxDatumsPerRequest	Numero massimo di riferimenti da includere in una richiesta ad Amazon CloudWatch	Il valore di default è 20, uguale a quello di default di Amazon CloudWatch

Supporto e miglioramenti delle configurazioni Amazon EMR Hudi

- I clienti possono ora sfruttare l'API di configurazione EMR e la funzione di riconfigurazione per configurare le configurazioni Hudi a livello di cluster. Un nuovo supporto per la configurazione basato su file è stato introdotto tramite `/etc/hudi/conf/hudi-defaults.conf` sulla linea di altre applicazioni come Spark, Hive ecc. EMR configura alcuni valori di default per migliorare l'esperienza utente:

— `hoodie.datasource.hive_sync.jdbcurl` è configurato per l'URL del server Hive del cluster e non deve più essere specificato. Ciò è particolarmente utile quando si esegue un

processo in modalità cluster Spark, dove in precedenza era necessario specificare l'IP principale Amazon EMR.

- Configurazioni specifiche di HBase, utili per l'utilizzo dell'indice HBase con Hudi.
- Configurazione specifica del provider di blocco Zookeeper, come discusso sotto il controllo della concorrenza, che semplifica l'utilizzo di Optimistic Concurrency Control (OCC).
- Sono state introdotte ulteriori modifiche per ridurre il numero di configurazioni che devi passare e per dedurre automaticamente laddove possibile:
 - La parola chiave `partitionBy` può essere utilizzata per specificare la colonna della partizione.
 - Quando si abilita Hive Sync, non è più obbligatorio passare `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Tali valori possono essere dedotti dal nome della tabella Hudi e dal campo della partizione.
 - `KEYGENERATOR_CLASS_OPT_KEY` non è obbligatorio passare e può essere dedotto da casi più semplici di `SimpleKeyGenerator` e `ComplexKeyGenerator`.

Avvertenze Hudi

- Hudi non supporta l'esecuzione vettorizzata nelle tabelle Hive for Merge on Read (MoR) e Bootstrap. Ad esempio: `count(*)` fallisce con la tabella in tempo reale Hudi quando `hive.vectorized.execution.enabled` è impostato su `VERO`. Come soluzione alternativa, puoi disabilitare la lettura vettorizzata impostando `hive.vectorized.execution.enabled` a `false`.
- Il supporto multi-writer non è compatibile con la caratteristica bootstrap Hudi.
- Flink Streamer e Flink SQL sono caratteristiche sperimentali di questa versione. Queste caratteristiche non sono consigliate per l'uso nelle distribuzioni di produzione.

Modifiche, miglioramenti e problemi risolti

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

- In precedenza, il riavvio manuale del gestore delle risorse su un cluster multi-master provocava il ricaricamento, da parte dei daemon su cluster di Amazon EMR come Zookeeper, di tutti i nodi

precedentemente disattivati o persi nel file `znode` di Zookeeper. Ciò causava il superamento dei limiti predefiniti in determinate situazioni. Ora Amazon EMR rimuove i record dei nodi disattivati o persi più vecchi di un'ora dal file Zookeeper e i limiti interni sono stati aumentati.

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Configurazione di un cluster per risolvere i problemi di prestazioni di Apache YARN Timeline Server versione 1 e 1.5

Le versioni 1 e 1.5 di Apache YARN Timeline Server possono causare problemi di prestazioni con cluster EMR molto attivi e di grandi dimensioni, in particolare con `yarn.resourcemanager.system-metrics-publisher.enabled=true`, che è l'impostazione predefinita in Amazon EMR. Un YARN Timeline Server v2 open source risolve il problema di prestazioni relativo alla scalabilità di YARN Timeline Server.

Altre soluzioni alternative per questo problema includono:

- La configurazione di `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.

- L'abilitazione della correzione per questo problema durante la creazione di un cluster, come descritto di seguito.

I seguenti rilasci di Amazon EMR contengono una correzione per questo problema di prestazioni di YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1 e 6.4.x

Per abilitare la correzione su una delle versioni di Amazon EMR sopra specificate, imposta queste proprietà su `true` in un file JSON di configurazioni che viene passato utilizzando il [parametro del comando `aws emr create-cluster`](#): `--configurations file:///./configurations.json`. Oppure abilita la correzione utilizzando l'[interfaccia utente della console di riconfigurazione](#).

Esempio del contenuto del file `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": []
}
```

- I server HttpFS e WebHDFS sono disabilitati per impostazione predefinita. Puoi abilitare nuovamente WebHDFS utilizzando la configurazione di Hadoop, `dfs.webhdfs.enabled`. Il server HttpFS può essere avviato utilizzando `sudo systemctl start hadoop-httpfs`.
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).
- Hive: le prestazioni delle query di scrittura sono migliorate abilitando l'uso di una directory Scratch su HDFS per l'ultimo processo. I dati temporanei per il processo finale vengono scritti su HDFS

anziché su Amazon S3 e le prestazioni sono migliorate perché i dati vengono spostati da HDFS alla posizione finale della tabella (Amazon S3) anziché tra i dispositivi Amazon S3.

- Hive: miglioramento del tempo di compilazione delle query fino a 2,5 volte con Glue metastore Partition Pruning.
- Per impostazione predefinita, quando le FDU integrate vengono passate da Hive al Hive Metastore Server, solo un sottoinsieme di tali FDU incorporati viene passato al Glue Metastore poiché Glue supporta solo operatori di espressioni limitate. Se hai impostato `hive.glue.partition.pruning.client=true`, quindi tutta la eliminazione delle partizioni avviene sul lato client. Se hai impostato `hive.glue.partition.pruning.server=true`, quindi tutta la eliminazione delle partizioni avviene sul lato server.

Problemi noti

- Le query Hue non funzionano in Amazon EMR 6.4.0 perché il server HttpFS di Apache Hadoop è disabilitato per impostazione predefinita. Per utilizzare Hue su Amazon EMR 6.4.0, avvia manualmente il server HttpFS sul nodo primario di Amazon EMR utilizzando `sudo systemctl start hadoop-httpfs` oppure [utilizza una fase di Amazon EMR](#).
- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.
- In Amazon EMR versione 6.4.0, Phoenix non supporta il componente dei connettori Phoenix.
- Per utilizzare le operazioni Spark con Apache Oozie, devi aggiungere la seguente configurazione al tuo file `workflow.xml` Oozie. In caso contrario, diverse librerie critiche come Hadoop e EMRFS non saranno presenti nella classpath degli esecutori Spark lanciati da Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.

- Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	3.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.3.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.47.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.13.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.13.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	3.2.1-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.2.1-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.4.4-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.4.4-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.4.4-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.4.4-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.4.4-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-5	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.2-amzn-5	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-5	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-5	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-5	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-5	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-5	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.8.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.8.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-trino	0.8.0-amzn-0	Libreria bundle per eseguire Trino con Hudi.
hudi-spark	0.8.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	5.1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.254.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.254.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.254.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
trino-coordinator	359	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i trino-workers.
trino-worker	359	Servizio per l'esecuzione di parti di una query.
trino-client	359	Trino client della riga di comando che viene installato su master di standby di un cluster HA dove server Trino non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing

Componente	Versione	Descrizione
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.1.2-amzn-0	Client a riga di comando Spark.
spark-history-server	3.1.2-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.1.2-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.1.2-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.4.1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.9.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.5.7	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.5.7	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.4.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
trino-log	Modifica i valori nel file log.properties di Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifica i valori nel file config.properties di Trino.	Restarts Trino-Server (for Trino)

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-password-authenticator	Modifica i valori nel file password-authenticator.properties di Trino.	Restarts Trino-Server (for Trino)
trino-env	Modifica i valori nel file trino-env.sh di Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifica i valori nel file node.properties di Trino.	Not available.
trino-connector-blackhole	Modifica i valori nel file blackhole.properties di Trino.	Not available.
trino-connector-cassandra	Modifica i valori nel file cassandra.properties di Trino.	Not available.
trino-connector-hive	Modifica i valori nel file hive.properties di Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifica i valori nel file jmx.properties di Trino.	Not available.
trino-connector-kafka	Modifica i valori nel file kafka.properties di Trino.	Not available.
trino-connector-localfile	Modifica i valori nel file localfile.properties di Trino.	Not available.
trino-connector-memory	Modifica i valori nel file memory.properties di Trino.	Not available.
trino-connector-mongodb	Modifica i valori nel file mongodb.properties di Trino.	Not available.
trino-connector-mysql	Modifica i valori nel file mysql.properties di Trino.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
trino-connector-postgresql	Modifica i valori nel file postgresql.properties di Trino.	Not available.
trino-connector-raptor	Modifica i valori nel file raptor.properties di Trino.	Not available.
trino-connector-redis	Modifica i valori nel file redis.properties di Trino.	Not available.
trino-connector-redshift	Modifica i valori nel file redshift.properties di Trino.	Not available.
trino-connector-tpch	Modifica i valori nel file tpch.properties di Trino.	Not available.
trino-connector-tpcds	Modifica i valori nel file tpcds.properties di Trino.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.3.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
SDK AWS per 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
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
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
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

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.

- Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	3.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.46.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.12.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.12.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-3.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-3.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-3.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-3.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	3.2.1-amzn-3.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-3.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-3.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-3.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.6-amzn-1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.6-amzn-1	Servizio per servire una o più regioni HBase.
hbase-client	2.2.6-amzn-1	Client di riga di comando HBase.
hbase-rest-server	2.2.6-amzn-1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.6-amzn-1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.2-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-4	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.7.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.7.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.7.0-amzn-0	Libreria bundle per eseguire PrestoSQL con Hudi.
hudi-spark	0.7.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.

Componente	Versione	Descrizione
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.2.2	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.245.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.245.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.245.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
prestosql-coordinator	350	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i prestosql-workers.
prestosql-worker	350	Servizio per l'esecuzione di parti di una query.
prestosql-client	350	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing

Componente	Versione	Descrizione
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.1.1-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.1.1-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.1.1-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.1.1-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.4.1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.9.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.3.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.

Classificazioni	Descrizione	Operazioni di riconfigurazione
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
presto-sql-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-env	Modifica i valori nel file <code>prestosql-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifica i valori nel file <code>node.properties</code> di PrestoSQL.	Not available.
prestosql-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di PrestoSQL.	Not available.
prestosql-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di PrestoSQL.	Not available.
prestosql-connector-hive	Modifica i valori nel file <code>hive.properties</code> di PrestoSQL.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di PrestoSQL.	Not available.
prestosql-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di PrestoSQL.	Not available.
prestosql-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di PrestoSQL.	Not available.
prestosql-connector-memory	Modifica i valori nel file <code>memory.properties</code> di PrestoSQL.	Not available.
prestosql-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di PrestoSQL.	Not available.
prestosql-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-postgresql	Modifica i valori nel file postgresql.properties di PrestoSQL.	Not available.
prestosql-connector-raptor	Modifica i valori nel file raptor.properties di PrestoSQL.	Not available.
prestosql-connector-redis	Modifica i valori nel file redis.properties di PrestoSQL.	Not available.
prestosql-connector-redshift	Modifica i valori nel file redshift.properties di PrestoSQL.	Not available.
prestosql-connector-tpch	Modifica i valori nel file tpch.properties di PrestoSQL.	Not available.
prestosql-connector-tpcds	Modifica i valori nel file tpcds.properties di PrestoSQL.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.3.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
SDK AWS per 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
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.3.0. Le modifiche sono relative alla versione 6.2.0.

Data del rilascio iniziale: 12 maggio 2021

Ultimo aggiornamento: 9 agosto 2021

Applicazioni supportate

- AWS SDK for Java versione 1.11.977
- CloudWatch Sink versione 2.1.0
- DynamoDB Connector versione 4.16.0
- EMRFS versione 2.46.0
- Amazon EMR Goodies versione 3.2.0
- Amazon EMR Kinesis Connector versione 3.5.0
- Amazon EMR Record Server versione 2.0.0
- Amazon EMR Scripts versione 2.5.0
- Flink versione 1.12.1
- Ganglia versione 3.7.2
- AWS Glue Hive Metastore Client versione 3.2.0
- Hadoop versione 3.2.1-amzn-3
- HBase versione 2.2.6-amzn-1
- HBase-operator-tools 1.0.0
- HCatalog versione 3.1.2-amzn-0
- Hive versione 3.1.2-amzn-4
- Hudi versione 0.7.0-amzn-0
- Hue versione 4.9.0
- Java JDK versione Corretto-8.282.08.1 (build 1.8.0_282-b08)
- JupyterHub versione 1.2.0
- Livy versione 0.7.0-incubating
- MXNet versione 1.7.0
- Oozie versione 5.2.1
- Phoenix versione 5.0.0
- Pig versione 0.17.0
- Presto versione 0.245.1-amzn-0
- PrestoSQL versione 350

- Apache Ranger KMS (crittografia trasparente multi-master) versione 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.1.0
- SageMaker Spark SDK versione 1.4.1
- Scala versione 2.12.10 (VM server OpenJDK a 64 bit, Java 1.8.0_282)
- Spark versione 3.1.1-amzn-0
- spark-rapids 0.4.1
- Sqoop versione 1.4.7
- TensorFlow versione 2.4.1
- tez versione 0.9.2
- Zeppelin versione 0.9.0
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.16.0

Nuove funzionalità di

- Amazon EMR supporta Amazon S3 Access Point, una funzionalità di Amazon S3 che consente di gestire facilmente l'accesso ai data lake condivisi. Utilizzando il tuo alias Amazon S3 Access Point, puoi semplificare l'accesso ai dati su larga scala su Amazon EMR. Puoi utilizzare gli Amazon S3 Access Point con tutte le versioni di Amazon EMR senza alcun costo aggiuntivo in tutte le Regioni AWS in cui è disponibile Amazon EMR. Per ulteriori informazioni sugli alias degli Access Point e degli Amazon S3 Access Point, consulta [Utilizzo di un alias in stile bucket per il punto di accesso](#) nella Guida per l'utente di Amazon S3.
- I nuovi parametri API di `DescribeReleaseLabel` e `ListReleaseLabel` forniscono i dettagli dell'etichetta di rilascio di Amazon EMR. L'utente può elencare a livello di programmazione i rilasci disponibili nella regione in cui viene eseguita la richiesta API ed elencare le applicazioni disponibili per un'etichetta di rilascio specifica di Amazon EMR. I parametri dell'etichetta di rilascio elencano anche i rilasci di Amazon EMR che supportano un'applicazione specifica, ad esempio Spark. Queste informazioni possono essere utilizzate per avviare a livello di programmazione cluster Amazon EMR. Ad esempio, è possibile avviare un cluster utilizzando la versione più recente dai risultati di `ListReleaseLabel`. Per ulteriori informazioni, consulta [DescribeReleaseLabel](#) e [ListReleaseLabels](#) nella Guida di riferimento alle API di Amazon EMR.
- Con Amazon EMR 6.3.0, è possibile avviare un cluster che si integra nativamente con Apache Ranger. Apache Ranger è un framework open source che consente di abilitare, monitorare e

gestire la sicurezza completa dei dati attraverso la piattaforma Hadoop. Per ulteriori informazioni, consulta [Apache Ranger](#). L'integrazione nativa consente di utilizzare Apache Ranger per imporre un controllo granulare di accesso ai dati su Amazon EMR. Consulta [Integrazione di Amazon EMR con Apache Ranger](#) nella Guida alla gestione di Amazon EMR.

- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).
- Stato del supporto IMDS (Instance Metadata Service) V2: per Amazon EMR 6.2 o versioni successive, i componenti Amazon EMR utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Se si disabilita IMDSv1 nelle versioni precedenti di Amazon EMR 6.x, si verificherà un errore di avvio del cluster.

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo

- era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
 - La modalità di spiegazione dell'interfaccia utente di Spark SQL è cambiata di default da `extended` a `formatted` in [Spark 3.1](#). Amazon EMR l'ha ripristinata a `extended` per includere le informazioni sul piano logico nell'interfaccia utente di Spark SQL. Questa può essere ripristinata impostando `spark.sql.ui.explainMode` su `formatted`.
 - I seguenti commit sono stati sottoposti al backport dal ramo principale di Spark.
 - [\[SPARK-34752\]](#)[BUILD] Spostamento di Jetty a 9.4.37 per l'indirizzo CVE-2020-27223.
 - [\[SPARK-34534\]](#) Correzione dell'ordine `blockIds` quando si utilizza `FetchShuffleBlocks` per recuperare i blocchi.
 - [\[SPARK-34681\]](#) [SQL] Correzione del bug per l'hash join esterno completo mescolato quando si costruisce il lato sinistro con condizione non uguale.
 - [\[SPARK-34497\]](#) [SQL] Correzione dei provider di connessione JDBC incorporati per ripristinare le modifiche al contesto di sicurezza JVM.
 - Per migliorare l'interoperabilità con il plug-in Nvidia Spark RAPIDS, è stata aggiunta una soluzione alternativa per risolvere un problema che impedisce l'attivazione dell'eliminazione dinamica delle partizioni quando si utilizza Nvidia Spark RAPIDS con l'esecuzione di query adattiva disabilitata; consulta il [problema RAPIDS n. 1378](#) e il [problema RAPIDS n. 1386](#). Per dettagli sulla nuova configurazione `spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse`, consulta il [problema RAPIDS n. 1386](#).
 - L'algoritmo predefinito del committer di output del file è stato modificato dall'algoritmo v2 all'algoritmo v1 in Spark 3.1 open source. Per ulteriori informazioni, consulta [Amazon EMR ottimizza le prestazioni di Spark: eliminazione dinamica delle partizioni](#).
 - Amazon EMR ha ripristinato l'algoritmo v2, il valore predefinito utilizzato nelle versioni di Amazon EMR precedenti alla 6.x, per evitare la regressione delle prestazioni. Per ripristinare il comportamento di Spark 3.1 open source, imposta `spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` su 1. Spark open source ha apportato questa modifica perché il task commit nell'algoritmo del committer v2

di output del file non è atomico, il che può causare un problema di correttezza dei dati di output in alcuni casi. Tuttavia, anche il task commit nell'algoritmo v1 non è atomico. In alcuni casi, il task commit include un'eliminazione eseguita prima di una rinomina. Ciò può causare un problema di correttezza dei dati silenti.

- Risolti i problemi di dimensionamento gestito nelle versioni precedenti di Amazon EMR e apportati miglioramenti in modo da ridurre significativamente le percentuali di errore delle applicazioni.
- Installato il AWS Java SDK Bundle su ogni nuovo cluster. Si tratta di un singolo jar contenente tutti gli SDK di servizio e le relative dipendenze, anziché i singoli jar di componenti. Per ulteriori informazioni, consulta [Dipendenza dell'SDK in bundle](#).

Problemi noti

- Per i cluster della sottorete privata di Amazon EMR 6.3.0 e 6.2.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio httpd sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.
- Quando AWS Glue Data Catalog è abilitato, l'utilizzo di Spark per accedere a un AWS Glue DB con l'URI di posizione della stringa null potrebbe non andare a buon fine. Questo accade alle versioni precedenti di Amazon EMR, ma SPARK-31709 (<https://issues.apache.org/jira/browse/SPARK-31709>) lo rende applicabile a più casi. Ad esempio, quando si crea una tabella all'interno dell'AWS Glue DB di default il cui URI di posizione è una stringa null, `spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` ha esito negativo con il messaggio "Cannot create a Path from an empty string (Impossibile creare un percorso da una stringa vuota)". Per aggirare questo problema, imposta manualmente un URI di posizione dei tuoi database AWS Glue, quindi crea tabelle all'interno di questi database utilizzando Spark.
- In Amazon EMR 6.3.0, PrestoSQL è stato aggiornato dalla versione 343 alla versione 350. Sono state apportate due modifiche relative alla protezione rispetto all'open source. Il controllo dell'accesso al catalogo basato su file viene modificato da deny a allow quando le regole delle proprietà di tabella, schema o sessione non sono definite. Inoltre, il controllo dell'accesso al sistema basato su file viene modificato per supportare i file senza regole di catalogo definite. In questo caso, è consentito l'intero accesso ai cataloghi.

Per ulteriori informazioni, consulta [Rilascio 344 \(9 ottobre 2020\)](#)

- Nota che la directory utente Hadoop (/home/hadoop) è leggibile da chiunque. Infatti, è dotata delle autorizzazioni di directory Unix 755 (drwxr-xr-x) per consentire l'accesso in lettura da framework quali Hive. È possibile inserire i file in /home/hadoop e nelle sue sottodirectory, ma occorre essere consapevoli delle autorizzazioni in vigore su tali directory per proteggere le informazioni sensibili.
- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica /etc/systemd/system/instance-controller.service per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Per utilizzare le operazioni Spark con Apache Oozie, devi aggiungere la seguente configurazione al tuo file `workflow.xml` Oozie. In caso contrario, diverse librerie critiche come Hadoop e EMRFS non saranno presenti nella classpath degli esecutori Spark lanciati da Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.1	SDK Spark Amazon SageMaker
emr-ddb	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.1.0	Connettore di EMR S3Select
emrfs	2.46.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.12.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.12.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
		istema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-3	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	3.2.1-amzn-3	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	3.2.1-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.6-amzn-1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.6-amzn-1	Servizio per servire una o più regioni HBase.
hbase-client	2.2.6-amzn-1	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	2.2.6-amzn-1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.6-amzn-1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-4	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hudi	0.7.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.7.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.7.0-amzn-0	Libreria bundle per eseguire PrestoSQL con Hudi.
hudi-spark	0.7.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.2.2	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.245.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.245.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.245.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
prestosql-coordinator	350	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i prestosql-workers.

Componente	Versione	Descrizione
prestoql-worker	350	Servizio per l'esecuzione di parti di una query.
prestoql-client	350	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.1.1-amzn-0	Client a riga di comando Spark.
spark-history-server	3.1.1-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.1.1-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.1.1-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.4.1	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.3.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.

Classificazioni	Descrizione	Operazioni di riconfigurazione
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
presto-sql-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-env	Modifica i valori nel file <code>prestosql-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifica i valori nel file <code>node.properties</code> di PrestoSQL.	Not available.
prestosql-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di PrestoSQL.	Not available.
prestosql-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di PrestoSQL.	Not available.
prestosql-connector-hive	Modifica i valori nel file <code>hive.properties</code> di PrestoSQL.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di PrestoSQL.	Not available.
prestosql-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di PrestoSQL.	Not available.
prestosql-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di PrestoSQL.	Not available.
prestosql-connector-memory	Modifica i valori nel file <code>memory.properties</code> di PrestoSQL.	Not available.
prestosql-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di PrestoSQL.	Not available.
prestosql-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-postgresql	Modifica i valori nel file postgresql.properties di PrestoSQL.	Not available.
prestosql-connector-raptor	Modifica i valori nel file raptor.properties di PrestoSQL.	Not available.
prestosql-connector-redis	Modifica i valori nel file redis.properties di PrestoSQL.	Not available.
prestosql-connector-redshift	Modifica i valori nel file redshift.properties di PrestoSQL.	Not available.
prestosql-connector-tpch	Modifica i valori nel file tpch.properties di PrestoSQL.	Not available.
prestosql-connector-tpcds	Modifica i valori nel file tpcds.properties di PrestoSQL.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.2.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
SDK AWS per 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
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
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
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

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.1	SDK Spark Amazon SageMaker
emr-ddb	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.0.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.16.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.0.0	Connettore di EMR S3Select
emrfs	2.44.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
		istema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-2.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-2.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-2.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-2.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-2.1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	3.2.1-amzn-2.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-2.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.6-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.6-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.2.6-amzn-0	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	2.2.6-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.6-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-3	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hudi	0.6.0-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.6.0-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.6.0-amzn-1	Libreria bundle per eseguire PrestoSQL con Hudi.
hudi-spark	0.6.0-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.8.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.238.3-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.238.3-amzn-1	Servizio per l'esecuzione di parti di una query.
presto-client	0.238.3-amzn-1	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
prestosql-coordinator	343	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i prestosql-workers.

Componente	Versione	Descrizione
prestosql-worker	343	Servizio per l'esecuzione di parti di una query.
prestosql-client	343	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.0.1-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.0.1-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.0.1-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.0.1-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.2.0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.3.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0-preview1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.2.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Not available.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Not available.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Not available.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
https-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
https-site	Modifica i valori nel file https-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
presto-sql-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-node	Modifica i valori nel file <code>node.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-hive	Modifica i valori nel file <code>hive.properties</code> di PrestoSQL.	Restarts Presto-Server (for PrestoSQL)
presto-sql-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-kafka	Modifica i valori nel file kafka.properties di PrestoSQL.	Not available.
prestosql-connector-localfile	Modifica i valori nel file localfile.properties di PrestoSQL.	Not available.
prestosql-connector-memory	Modifica i valori nel file memory.properties di PrestoSQL.	Not available.
prestosql-connector-mongodb	Modifica i valori nel file mongodb.properties di PrestoSQL.	Not available.
prestosql-connector-mysql	Modifica i valori nel file mysql.properties di PrestoSQL.	Not available.
prestosql-connector-postgresql	Modifica i valori nel file postgresql.properties di PrestoSQL.	Not available.
prestosql-connector-raptor	Modifica i valori nel file raptor.properties di PrestoSQL.	Not available.
prestosql-connector-redis	Modifica i valori nel file redis.properties di PrestoSQL.	Not available.
prestosql-connector-redshift	Modifica i valori nel file redshift.properties di PrestoSQL.	Not available.
prestosql-connector-tpch	Modifica i valori nel file tpch.properties di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-tpcds	Modifica i valori nel file tpcds.properties di PrestoSQL.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.2.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
SDK AWS per 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
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	-

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-
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	-

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	338	338	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.2.0. Le modifiche sono relative alla versione 6.1.0.

Data del rilascio iniziale: 9 dicembre 2020

Ultimo aggiornamento: 4 ottobre 2021

Applicazioni supportate

- AWS SDK for Java versione 1.11.828
- emr-record-server versione 1.7.0
- Flink versione 1.11.2
- Ganglia versione 3.7.2
- Hadoop versione 3.2.1-amzn-1
- HBase versione 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog versione 3.1.2-amzn-0
- Hive versione 3.1.2-amzn-3
- Hudi versione 0.6.0-amzn-1
- Hue versione 4.8.0
- JupyterHub versione 1.1.0

- Livy versione 0.7.0
- MXNet versione 1.7.0
- Oozie versione 5.2.0
- Phoenix versione 5.0.0
- Pig versione 0.17.0
- Presto versione 0.238.3-amzn-1
- PrestoSQL versione 343
- Spark versione 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow versione 2.3.1
- Zeppelin versione 0.9.0-preview1
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.16.0

Nuove funzionalità di

- HBase: rimossa la rinomina in fase di commit e aggiunto il tracciamento HFile persistente. Consulta [Tracciamento HFile persistente](#) nella Guida ai rilasci di Amazon EMR.
- HBase: backport per [Creazione di una configurazione che costringe a memorizzare nella cache i blocchi sulla compattazione](#).
- PrestoDB: miglioramenti all'eliminazione delle partizioni dinamiche. Join Reorder basato su regole funziona su dati non partizionati.
- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).
- Stato del supporto IMDS (Instance Metadata Service) V2: per Amazon EMR 6.2 o versioni successive, i componenti Amazon EMR utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Se si disabilita IMDSv1 nelle versioni precedenti di Amazon EMR 6.x, si verificherà un errore di avvio del cluster.

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Spark: miglioramenti delle prestazioni del runtime Spark.

Problemi noti

- Amazon EMR 6.2 dispone di autorizzazioni errate impostate sul file `/etc/cron.d/libinstance-controller-java` in EMR 6.2.0. Le autorizzazioni sul file sono 645 (-rw-r--r-x), quando dovrebbero essere 644 (-rw-r--r--). Di conseguenza, Amazon EMR versione 6.2 non registra i log dello stato

delle istanze e la directory `/emr/instance-logs` è vuota. Questo problema è stato risolto in Amazon EMR 6.3.0 e successive.

Per risolvere il problema, esegui il seguente script come operazione di bootstrap all'avvio del cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Per i cluster della sottorete privata di Amazon EMR 6.2.0 e 6.3.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio `httpd` sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.
- C'è un problema in Amazon EMR 6.2.0 in cui `httpd` ha continuamente esito negativo, causando la non disponibilità di Ganglia. Viene visualizzato un errore "cannot connect to the server (impossibile connettersi al server)". Per risolvere questo problema in un cluster che è già in esecuzione, abilita SSH sul nodo primario del cluster e aggiungi la riga `Listen 80` al file `httpd.conf` presente in `/etc/httpd/conf/httpd.conf`. Questo problema è stato risolto in Amazon EMR 6.3.0.
- HTTPD ha esito negativo su cluster EMR 6.2.0 quando si utilizza una configurazione di sicurezza. Ciò rende l'interfaccia utente dell'applicazione Web Ganglia non disponibile. Per accedere all'interfaccia utente dell'applicazione Web Ganglia, aggiungi `Listen 80` al file `/etc/httpd/conf/httpd.conf` sul nodo primario del cluster. Per ulteriori informazioni su come connettersi al cluster, consulta la sezione [Connect to the Primary Node Using SSH](#) (Connessione al nodo primario tramite SSH).

Inoltre, EMR Notebooks non riesce a stabilire una connessione con cluster EMR 6.2.0 quando si utilizza una configurazione di sicurezza. Il notebook non riuscirà a elencare i kernel e a inviare processi Spark. Ti consigliamo invece di utilizzare EMR Notebooks con un'altra versione di Amazon EMR.

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` ed `emr-6.2.0` si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon

EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
```

```
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

⚠ Important

Amazon EMR 6.1.0 e 6.2.0 includono un problema di prestazioni che può influire in modo critico su tutte le operazioni di inserimento, upsert ed eliminazione di Hudi. Se intendi utilizzare Hudi con Amazon EMR 6.1.0 o 6.2.0, contatta il supporto AWS per ottenere un RPM Hudi con patch.

⚠ Important

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Gli artifact di Amazon EMR 6.2.0 Maven non vengono pubblicati. Saranno pubblicati con una futura versione di Amazon EMR.
- Il tracciamento HFile persistente che utilizza la tabella di sistema Storefile di HBase non supporta la funzionalità di replica della regione HBase. Per ulteriori informazioni sulla replica della regione HBase, consulta [Lecture ad alta disponibilità coerenti con la tempistica](#).
- Differenze di versione dei bucket Hive tra Amazon EMR 6.x e EMR 5.x

EMR 5.x utilizza OOS Apache Hive 2, mentre EMR 6.x utilizza OOS Apache Hive 3. L'Hive2 open source utilizza bucket di versione 1, mentre l'Hive3 open source utilizza bucket di versione 2. Questa differenza di versione dei bucket tra Hive 2 (EMR 5.x) e Hive 3 (EMR 6.x) presuppone un funzionamento diverso dell'hashing dei bucket Hive. Vedere l'esempio sottostante.

La tabella seguente è un esempio creato rispettivamente in EMR 6.x ed EMR 5.x.

```
-- 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/';
```

Sono stati inseriti gli stessi dati in EMR 6.x ed EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Osservando la posizione S3, si nota che il nome del file di bucket è diverso, poiché la funzione di hashing è diversa tra EMR 6.x (Hive 3) e EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

Puoi notare la differenza di versione anche eseguendo il comando riportato di seguito nella CLI Hive in EMR 6.x. Potrai notare che la CLI restituisce bucket di versione 2.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono

altri 14 caratteri non di controllo: !"#%&'()*+,-. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.1.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-notebook-env	1.0.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.16.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.0.0	Connettore di EMR S3Select
emrfs	2.44.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.2.1-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-2	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.2.1-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.6-amzn-0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.6-amzn-0	Servizio per servire una o più regioni HBase.
hbase-client	2.2.6-amzn-0	Client di riga di comando HBase.
hbase-rest-server	2.2.6-amzn-0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.6-amzn-0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	3.1.2-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-3	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.6.0-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.6.0-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.6.0-amzn-1	Libreria bundle per eseguire PrestoSQL con Hudi.
hudi-spark	0.6.0-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.8.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.238.3-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.238.3-amzn-1	Servizio per l'esecuzione di parti di una query.
presto-client	0.238.3-amzn-1	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
prestoql-coordinator	343	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti prestoql-workers.
prestoql-worker	343	Servizio per l'esecuzione di parti di una query.
prestoql-client	343	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	3.0.1-amzn-0	Client a riga di comando Spark.
spark-history-server	3.0.1-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.0.1-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.0.1-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
spark-rapids	0.2.0	Plugin Nvidia Spark RAPIDS che accelera Apache Spark con GPU.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.3.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0-preview1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-6.2.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Not available.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Not available.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Not available.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, HiveMetaStore, and MapReduceHistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	This classification should not be reconfigured.
hdfs-env	Modifica i valori nell'ambiente HDFS.	Restarts Hadoop HDFS ZKFC.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat server.
hive	Impostazioni Amazon EMR per Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Not available.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
https-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
https-site	Modifica i valori nel file https-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.	Not available.
presto-sql-log	Modifica i valori nel file <code>log.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-config	Modifica i valori nel file <code>config.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server (for PrestoSQL)
presto-sql-node	Modifica i valori nel file <code>node.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di PrestoSQL.	Not available.
presto-sql-connector-hive	Modifica i valori nel file <code>hive.properties</code> di PrestoSQL.	Restarts Presto-Server (for PrestoSQL)
presto-sql-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-kafka	Modifica i valori nel file kafka.properties di PrestoSQL.	Not available.
prestosql-connector-localfile	Modifica i valori nel file localfile.properties di PrestoSQL.	Not available.
prestosql-connector-memory	Modifica i valori nel file memory.properties di PrestoSQL.	Not available.
prestosql-connector-mongodb	Modifica i valori nel file mongodb.properties di PrestoSQL.	Not available.
prestosql-connector-mysql	Modifica i valori nel file mysql.properties di PrestoSQL.	Not available.
prestosql-connector-postgresql	Modifica i valori nel file postgresql.properties di PrestoSQL.	Not available.
prestosql-connector-raptor	Modifica i valori nel file raptor.properties di PrestoSQL.	Not available.
prestosql-connector-redis	Modifica i valori nel file redis.properties di PrestoSQL.	Not available.
prestosql-connector-redshift	Modifica i valori nel file redshift.properties di PrestoSQL.	Not available.
prestosql-connector-tpch	Modifica i valori nel file tpch.properties di PrestoSQL.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
prestosql-connector-tpcds	Modifica i valori nel file tpcds.properties di PrestoSQL.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restart Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 6.1.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
SDK AWS per 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
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo

era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.

- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.3.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	3.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.0.0	Connettore di EMR S3Select
emrfs	2.42.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-1.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	3.2.1-amzn-1.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-1.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-1.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	3.2.1-amzn-1.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-1.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-1.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.5	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.5	Servizio per servire una o più regioni HBase.
hbase-client	2.2.5	Client di riga di comando HBase.
hbase-rest-server	2.2.5	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.5	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	3.1.2-amzn-2	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.2-incubating-amzn-2	Framework di elaborazione incrementale per alimentare la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.2-incubating-amzn-2	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Libreria bundle per eseguire PrestoSQL con Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.7.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.6.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.232	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.232	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
presto-client	0.232	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
prestoql-coordinator	338	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti prestoql-workers.
prestoql-worker	338	Servizio per l'esecuzione di parti di una query.
prestoql-client	338	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.0.0-amzn-0.1	Client a riga di comando Spark.
spark-history-server	3.0.0-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	3.0.0-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.0.0-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.1.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0-preview1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-6.1.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>hadoop-log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>hadoop-ssl-server</code> .

Classificazioni	Descrizione
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-env	Modifica i valori nell'ambiente HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive	Impostazioni Amazon EMR per Apache Hive.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.

Classificazioni	Descrizione
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.

Classificazioni	Descrizione
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.

Classificazioni	Descrizione
prestosql-log	Modifica i valori nel file log.properties di Presto.
prestosql-config	Modifica i valori nel file config.properties di Presto.
prestosql-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
prestosql-env	Modifica i valori nel file presto-env.sh di Presto.
prestosql-node	Modifica i valori nel file node.properties di PrestoSQL.
prestosql-connector-blackhole	Modifica i valori nel file blackhole.properties di PrestoSQL.
prestosql-connector-cassandra	Modifica i valori nel file cassandra.properties di PrestoSQL.
prestosql-connector-hive	Modifica i valori nel file hive.properties di PrestoSQL.
prestosql-connector-jmx	Modifica i valori nel file jmx.properties di PrestoSQL.
prestosql-connector-kafka	Modifica i valori nel file kafka.properties di PrestoSQL.
prestosql-connector-localfile	Modifica i valori nel file localfile.properties di PrestoSQL.
prestosql-connector-memory	Modifica i valori nel file memory.properties di PrestoSQL.
prestosql-connector-mongodb	Modifica i valori nel file mongodb.properties di PrestoSQL.

Classificazioni	Descrizione
prestosql-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di PrestoSQL.
prestosql-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di PrestoSQL.
prestosql-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di PrestoSQL.
prestosql-connector-redis	Modifica i valori nel file <code>redis.properties</code> di PrestoSQL.
prestosql-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di PrestoSQL.
prestosql-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di PrestoSQL.
prestosql-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di PrestoSQL.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 6.1.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
SDK AWS per 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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.1.0. Le modifiche sono relative alla versione 6.0.0.

Data del rilascio iniziale: 4 settembre 2020

Ultimo aggiornamento: 15 ottobre 2020

Applicazioni supportate

- AWS SDK for Java versione 1.11.828
- Flink versione 1.11.0
- Ganglia versione 3.7.2
- Hadoop versione 3.2.1-amzn-1
- HBase versione 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog versione 3.1.2-amzn-0
- Hive versione 3.1.2-amzn-1
- Hudi versione 0.5.2 in fase di sviluppo
- Hue versione 4.7.1
- JupyterHub versione 1.1.0
- Livy versione 0.7.0
- MXNet versione 1.6.0
- Oozie versione 5.2.0
- Phoenix versione 5.0.0
- Presto versione 0.232

- PrestoSQL versione 338
- Spark versione 3.0.0-amzn-0
- TensorFlow versione 2.1.0
- Zeppelin versione 0.9.0-preview1
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.14.0

Nuove funzionalità di

- I tipi di istanze ARM sono supportati a partire da Amazon EMR versione 5.30.0 e Amazon EMR versione 6.1.0.
- I tipi di istanze per uso generico M6g sono supportati a partire dalle versioni di Amazon EMR 6.1.0 e 5.30.0. Per ulteriori informazioni, consulta [Tipi di istanza supportati](#) nella Guida alla gestione di Amazon EMR.
- La caratteristica del gruppo di collocamento EC2 è supportata a partire da Amazon EMR versione 5.23.0 come opzione per i cluster con più nodi primari. Attualmente, solo i tipi di nodi primari sono supportati dalla caratteristica del gruppo di collocamento e la strategia SPREAD viene applicata a tali nodi primari. La strategia SPREAD colloca un piccolo gruppo di istanze su hardware sottostante separato per evitare la perdita di più nodi primari in caso di guasto hardware. Per ulteriori informazioni, consulta [Integrazione di EMR con il gruppo di collocamento EC2](#) nella Guida alla gestione di Amazon EMR.
- Scalabilità gestita: con la versione 6.1.0 di Amazon EMR, puoi abilitare Scalabilità gestita di Amazon EMR per aumentare o diminuire automaticamente il numero di istanze o unità nel cluster in base al carico di lavoro. Amazon EMR valuta continuamente i parametri dei cluster per prendere decisioni di dimensionamento che ottimizzano i cluster in termini di costi e velocità. Il dimensionamento gestito è disponibile anche su Amazon EMR versione 5.30.0 e successive, tranne 6.0.0. Per maggiori informazioni, consulta [Dimensionamento delle risorse del cluster](#) nella Guida alla gestione di Amazon EMR.
- PrestoSQL versione 338 è supportata con EMR 6.1.0. Per ulteriori informazioni, consulta [Presto](#).
 - PrestoSQL è supportata solo su EMR 6.1.0 e versioni successive, non su EMR 6.0.0 o EMR 5.x.
 - Il nome dell'applicazione Presto continua ad essere utilizzato per installare PrestoDB sui cluster. Per installare PrestoSQL sui cluster, utilizza il nome dell'applicazione PrestoSQL.
 - È possibile installare PrestoDB o PrestoSQL, ma non è possibile installare entrambi in un singolo cluster. Se durante il tentativo di creare un cluster vengono specificati sia PrestoDB

che PrestoSQL, si verifica un errore di convalida e la richiesta di creazione del cluster ha esito negativo.

- PrestoSQL è supportato su entrambi i cluster single-master e multi-master. Nei cluster multi-master, è necessario un metastore Hive esterno per eseguire PrestoSQL o PrestoDB. Consulta la sezione [Applicazioni supportate in un cluster EMR con più nodi primari](#).
- Supporto per l'autenticazione automatica ECR su Apache Hadoop e Apache Spark con Docker: gli utenti di Docker possono utilizzare le immagini Docker da Docker Hub e Amazon Elastic Container Registry (Amazon ECR) per definire le dipendenze dell'ambiente e della libreria.

Consulta [Configurazione di Docker](#) ed [Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.x](#).

- EMR supporta le transazioni Apache Hive ACID: Amazon EMR 6.1.0 aggiunge il supporto per le transazioni Hive ACID in modo che sia conforme alle proprietà ACID di un database. Con questa funzione, puoi eseguire operazioni INSERT, UPDATE, DELETE, e MERGE nelle tabelle gestite da Hive con dati in Amazon Simple Storage Service (Amazon S3). Questa è una caratteristica chiave per casi d'uso come l'importazione dati in streaming, il riavvio dei dati, gli aggiornamenti in blocco tramite MERGE e la graduale variazione delle dimensioni. Per ulteriori informazioni, inclusi esempi di configurazione e casi d'uso, consulta [Amazon EMR supporta le transazioni Apache Hive ACID](#).

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.

- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Apache Flink non è supportata su EMR 6.0.0, ma è supportata su EMR 6.1.0 con Flink 1.11.0. Si tratta della prima versione di Flink a supportare ufficialmente Hadoop 3. Consulta [Annuncio del rilascio di Apache Flink 1.11.0](#).
- Ganglia è stato rimosso dai pacchetti EMR 6.1.0 di default.

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 e 6.2.0 includono un problema di prestazioni che può influire in modo critico su tutte le operazioni di inserimento, upsert ed eliminazione di Hudi. Se intendi utilizzare Hudi con Amazon EMR 6.1.0 o 6.2.0, contatta il supporto AWS per ottenere un RPM Hudi con patch.

- L'impostazione di configurazioni di garbage collection personalizzate con `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions` genera un errore di avvio del driver o dell'executor con EMR 6.1 a causa di una configurazione di garbage collection in conflitto. Con il rilascio EMR 6.1.0, è invece necessario specificare la configurazione di garbage collection Spark personalizzata per driver ed executor con le proprietà `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Per maggiori informazioni, consulta [Ambiente di runtime Apache Spark](#) e [Configurazione di garbage collection Spark su Amazon EMR 6.1.0](#).
- L'utilizzo di Pig con Oozie (e all'interno di Hue, dal momento che Hue utilizza le operazioni Oozie per eseguire gli script Pig) genera un errore che impedisce il caricamento di una libreria native-izo. Questo messaggio di errore è informativo e non impedisce l'esecuzione di Pig.
- Supporto della simultaneità di Hudi: attualmente, le scritture simultanee in una singola tabella Hudi non sono supportate. Inoltre, Hudi esegue il rollback di tutte le modifiche apportate dalle istanze di scrittura in corso prima di consentire l'avvio di una nuova istanza di scrittura. Le scritture simultanee possono interferire con questo meccanismo e introdurre condizioni di competizione, le quali possono causare il danneggiamento dei dati. È necessario assicurarsi che, come parte del flusso di lavoro di elaborazione dei dati, ci sia sempre una sola istanza di scrittura Hudi che opera su una tabella Hudi. Hudi supporta più istanze di lettura simultanee che operano sulla stessa tabella Hudi.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Si è verificato un problema in Amazon EMR 6.1.0 che interessa i cluster che eseguono Presto. Dopo un lungo periodo di tempo (giorni), il cluster potrebbe generare errori come "su: failed to execute /bin/bash: Resource temporarily unavailable (su: impossibile eseguire /bin/bash: risorsa temporaneamente non disponibile)" o "shell request failed on channel 0 (richiesta shell non riuscita sul canale 0)". Questo problema è causato da un processo interno di Amazon EMR (InstanceController) che genera un numero eccessivo di processi leggeri (LWP) e, alla fine, fa sì che l'utente Hadoop superi il limite `nproc`. Ciò impedisce all'utente di aprire ulteriori processi. La soluzione a questo problema consiste nell'eseguire l'aggiornamento a EMR 6.2.0.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.3.0	SDK Spark Amazon SageMaker
emr-ddb	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	2.0.0	Connettore di EMR S3Select
emrfs	2.42.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	3.2.1-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	3.2.1-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.5	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.5	Servizio per servire una o più regioni HBase.
hbase-client	2.2.5	Client di riga di comando HBase.
hbase-rest-server	2.2.5	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.5	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	3.1.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-2	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.2-incubating-amzn-2	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.2-incubating-amzn-2	Libreria bundle per eseguire Presto con Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Libreria bundle per eseguire PrestoSQL con Hudi.

Componente	Versione	Descrizione
hudi-spark	0.5.2-incubating-amzn-2	Libreria bundle per eseguire Spark con Hudi.
hue-server	4.7.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.6.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.232	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.232	Servizio per l'esecuzione di parti di una query.
presto-client	0.232	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
prestosql-coordinator	338	Servizio per l'accettazione di query e la gestione dell'esecuzione di query tra componenti i prestosql-workers.
prestosql-worker	338	Servizio per l'esecuzione di parti di una query.
prestosql-client	338	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing

Componente	Versione	Descrizione
ranger-kms-server	2.0.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	3.0.0-amzn-0	Client a riga di comando Spark.
spark-history-server	3.0.0-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	3.0.0-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	3.0.0-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.1.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0-preview1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-6.1.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-env	Modifica i valori nell'ambiente HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive	Impostazioni Amazon EMR per Apache Hive.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2

Classificazioni	Descrizione
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.

Classificazioni	Descrizione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.

Classificazioni	Descrizione
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
prestosql-log	Modifica i valori nel file log.properties di Presto.
prestosql-config	Modifica i valori nel file config.properties di Presto.
prestosql-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
prestosql-env	Modifica i valori nel file presto-env.sh di Presto.
prestosql-node	Modifica i valori nel file node.properties di PrestoSQL.
prestosql-connector-blackhole	Modifica i valori nel file blackhole.properties di PrestoSQL.
prestosql-connector-cassandra	Modifica i valori nel file cassandra.properties di PrestoSQL.
prestosql-connector-hive	Modifica i valori nel file hive.properties di PrestoSQL.
prestosql-connector-jmx	Modifica i valori nel file jmx.properties di PrestoSQL.
prestosql-connector-kafka	Modifica i valori nel file kafka.properties di PrestoSQL.

Classificazioni	Descrizione
prestosql-connector-localfile	Modifica i valori nel file localfile.properties di PrestoSQL.
prestosql-connector-memory	Modifica i valori nel file memory.properties di PrestoSQL.
prestosql-connector-mongodb	Modifica i valori nel file mongodb.properties di PrestoSQL.
prestosql-connector-mysql	Modifica i valori nel file mysql.properties di PrestoSQL.
prestosql-connector-postgresql	Modifica i valori nel file postgresql.properties di PrestoSQL.
prestosql-connector-raptor	Modifica i valori nel file raptor.properties di PrestoSQL.
prestosql-connector-redis	Modifica i valori nel file redis.properties di PrestoSQL.
prestosql-connector-redshift	Modifica i valori nel file redshift.properties di PrestoSQL.
prestosql-connector-tpch	Modifica i valori nel file tpch.properties di PrestoSQL.
prestosql-connector-tpcds	Modifica i valori nel file tpcds.properties di PrestoSQL.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.

Classificazioni	Descrizione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 6.0.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questa versione sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
SDK AWS per 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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Scala	2.12.10	2.12.10	2.12.10	2.11.12
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.

- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.6	SDK Spark Amazon SageMaker
emr-ddb	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	3.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.5.0	Connettore di EMR S3Select
emrfs	2.39.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	3.2.1-amzn-0.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-0.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-0.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-0.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-0.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-0.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-0.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.3	Servizio per servire una o più regioni HBase.
hbase-client	2.2.3	Client di riga di comando HBase.
hbase-rest-server	2.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	3.1.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	3.1.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-0	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.0-incubating-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.0-incubating-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.230	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.230	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
presto-client	0.230	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
r	3.4.3	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.
zeppelin-server	0.9.0-SNAPSHOT	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-6.0.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.

Classificazioni	Descrizione
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-env	Modifica i valori nell'ambiente HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive	Impostazioni Amazon EMR per Apache Hive.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.

Classificazioni	Descrizione
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.

Classificazioni	Descrizione
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 6.0.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questa versione sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
SDK AWS per Java	1.11.828	1.11.828	1.11.711	1.11.711

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.0.0.

Data del rilascio iniziale: 10 marzo 2020

Applicazioni supportate

- AWS SDK for Java versione 1.11.711
- Ganglia versione 3.7.2
- Hadoop versione 3.2.1
- HBase versione 2.2.3
- HCatalog versione 3.1.2
- Hive versione 3.1.2
- Hudi versione 0.5.0 in fase di sviluppo
- Hue versione 4.4.0
- JupyterHub versione 1.0.0

- Livy versione 0.6.0
- MXNet versione 1.5.1
- Oozie versione 5.1.0
- Phoenix versione 5.0.0
- Presto versione 0.230
- Spark versione 2.4.4
- TensorFlow versione 1.14.0
- Zeppelin versione 0.9.0-SNAPSHOT
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.14.0

Note

Flink, Sqoop, Pig e Mahout non sono disponibili in Amazon EMR versione 6.0.0.

Nuove funzionalità di

- Supporto runtime docker YARN – Applicazioni YARN, come i lavori Spark, ora possono essere eseguite nel contesto di un container Docker. Ciò consente di definire facilmente le dipendenze in un'immagine Docker senza la necessità di installare librerie personalizzate nel cluster Amazon EMR. Per ulteriori informazioni, consulta [Configurazione dell'integrazione Docker](#) ed [Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.0.0](#).
- Supporto LLAP Hive - Hive ora supporta la modalità di esecuzione LLAP per migliorare le prestazioni delle query. Per ulteriori informazioni, consulta la sezione relativa all'[utilizzo di LLAP Hive](#).

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e

dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.

- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Amazon Linux
 - Amazon Linux 2 è il sistema operativo per la serie EMR 6.x.
 - `systemd` viene utilizzato per la gestione dei servizi al posto di `upstart` utilizzato in Amazon Linux 1.
- Java Development Kit (JDK)
 - Corretto JDK 8 è il JDK predefinito per la serie delle versioni EMR 6.x.
- Scala
 - Scala 2.12 è usato con Apache Spark e Apache Livy.
- Python 3
 - Python 3 è ora la versione predefinita di Python in EMR.
- Etichette nodo YARN
 - A partire dalla serie di rilascio Amazon EMR 6.x, la funzione etichette nodo YARN è disabilitata per impostazione predefinita. Per impostazione predefinita, i processi master dell'applicazione

possono essere eseguiti su entrambi i nodi di task e core. È possibile abilitare la caratteristica etichette nodo YARN configurando le seguenti proprietà: `yarn.node-labels.enabled` e `yarn.node-labels.am.default-node-label-expression`. Per ulteriori informazioni, consulta la sezione [Understanding Primary, Core, and Task Nodes](#) (Informazioni sui nodi primari, core e attività).

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` ed `emr-6.2.0` si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione `ulimit` di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione `ulimit` durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'`ulimit` del controller dell'istanza su un massimo di 65536 file.

Impostazione di un `ulimit` esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- La shell interattiva di Spark, inclusi PySpark, SparkR e spark-shell, non supporta l'utilizzo di Docker con librerie aggiuntive.
- Per utilizzare Python 3 con Amazon EMR versione 6.0.0, è necessario aggiungere PATH a `yarn.nodemanager.env-whitelist`.
- La funzionalità Live Long and Process (LLAP) non è supportata quando si utilizza AWS Glue Data Catalog come metastore per Hive.
- Quando si utilizza Amazon EMR 6.0.0 con l'integrazione di Spark e Docker, è necessario configurare le istanze nel cluster con lo stesso tipo di istanza e la stessa quantità di volumi EBS per evitare errori durante l'invio di un processo Spark con il runtime di Docker.
- In Amazon EMR 6.0.0, la modalità di archiviazione di HBase su Amazon S3 è influenzata dal problema [HBASE-24286](#). Impossibile inizializzare il master HBase quando il cluster viene creato utilizzando dati S3 esistenti.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.6	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	3.0.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.5.0	Connettore di EMR S3Select
<code>emrfs</code>	2.39.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	3.2.1-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	3.2.1-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	3.2.1-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	3.2.1-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	3.2.1-amzn-0	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	3.2.1-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	3.2.1-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	3.2.1-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	3.2.1-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	3.2.1-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	2.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	2.2.3	Servizio per servire una o più regioni HBase.
hbase-client	2.2.3	Client di riga di comando HBase.
hbase-rest-server	2.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	2.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	3.1.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	3.1.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	3.1.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	3.1.2-amzn-0	Client a riga di comando Hive.
hive-hbase	3.1.2-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	3.1.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.0-incubating-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.0-incubating-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MariaDB.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Le librerie Phoenix per server e client
phoenix-query-server	5.0.0-HBase-2.0	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.230	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.230	Servizio per l'esecuzione di parti di una query.
presto-client	0.230	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
r	3.4.3	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.41+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.9.0-SNAPSHOT	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-6.0.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-env	Modifica i valori nell'ambiente HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive	Impostazioni Amazon EMR per Apache Hive.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.

Classificazioni	Descrizione
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.

Classificazioni	Descrizione
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.

Classificazioni	Descrizione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Versioni del rilascio di Amazon EMR 5.x

Questa sezione contiene le versioni dell'applicazione, le note di rilascio, le versioni dei componenti e le classificazioni di configurazione disponibili in ogni versione di rilascio di Amazon EMR 5x.

Quando si avvia un cluster, è possibile scegliere tra diversi rilasci di Amazon EMR. In questo modo l'utente può testare e utilizzare le versioni dell'applicazione che soddisfano i propri requisiti di compatibilità. È possibile specificare la versione di rilascio con l'etichetta del rilascio. Le etichette di rilascio sono sotto forma di `emr-x.x.x`. For example, `emr-6.14.0`.

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Per una tabella completa delle versioni delle applicazioni in ogni versione di Amazon EMR 5.x, consulta [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#).

Argomenti

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versione 5.36.1 di Amazon EMR](#)
- [Rilascio 5.36.0 di Amazon EMR](#)
- [Amazon EMR versione 5.35.0](#)
- [Rilascio 5.34.0 di Amazon EMR](#)
- [Rilascio 5.33.1 di Amazon EMR](#)
- [Rilascio 5.33.0 di Amazon EMR](#)

- [Rilascio 5.32.1 di Amazon EMR](#)
- [Rilascio 5.32.0 di Amazon EMR](#)
- [Rilascio 5.31.1 di Amazon EMR](#)
- [Rilascio 5.31.0 di Amazon EMR](#)
- [Rilascio 5.30.2 di Amazon EMR](#)
- [Rilascio 5.30.1 di Amazon EMR](#)
- [Rilascio 5.30.0 di Amazon EMR](#)
- [Rilascio 5.29.0 di Amazon EMR](#)
- [Rilascio 5.28.1 di Amazon EMR](#)
- [Rilascio 5.28.0 di Amazon EMR](#)
- [Rilascio 5.27.1 di Amazon EMR](#)
- [Rilascio 5.27.0 di Amazon EMR](#)
- [Rilascio 5.26.0 di Amazon EMR](#)
- [Rilascio 5.25.0 di Amazon EMR](#)
- [Rilascio 5.24.1 di Amazon EMR](#)
- [Rilascio 5.24.0 di Amazon EMR](#)
- [Rilascio 5.23.1 di Amazon EMR](#)
- [Rilascio 5.23.0 di Amazon EMR](#)
- [Rilascio 5.22.0 di Amazon EMR](#)
- [Rilascio 5.21.2 di Amazon EMR](#)
- [Rilascio 5.21.1 di Amazon EMR](#)
- [Rilascio 5.21.0 di Amazon EMR](#)
- [Rilascio 5.20.1 di Amazon EMR](#)
- [Rilascio 5.20.0 di Amazon EMR](#)
- [Rilascio 5.19.1 di Amazon EMR](#)
- [Rilascio 5.19.0 di Amazon EMR](#)
- [Rilascio 5.18.1 di Amazon EMR](#)
- [Rilascio 5.18.0 di Amazon EMR](#)
- [Rilascio 5.17.2 di Amazon EMR](#)
- [Rilascio 5.17.1 di Amazon EMR](#)

- [Rilascio 5.17.0 di Amazon EMR](#)
- [Rilascio 5.16.1 di Amazon EMR](#)
- [Rilascio 5.16.0 di Amazon EMR](#)
- [Rilascio 5.15.1 di Amazon EMR](#)
- [Rilascio 5.15.0 di Amazon EMR](#)
- [Rilascio 5.14.2 di Amazon EMR](#)
- [Rilascio 5.14.1 di Amazon EMR](#)
- [Rilascio 5.14.0 di Amazon EMR](#)
- [Rilascio 5.13.1 di Amazon EMR](#)
- [Rilascio 5.13.0 di Amazon EMR](#)
- [Rilascio 5.12.3 di Amazon EMR](#)
- [Rilascio 5.12.2 di Amazon EMR](#)
- [Rilascio 5.12.1 di Amazon EMR](#)
- [Rilascio 5.12.0 di Amazon EMR](#)
- [Rilascio 5.11.4 di Amazon EMR](#)
- [Rilascio 5.11.3 di Amazon EMR](#)
- [Rilascio 5.11.2 di Amazon EMR](#)
- [Rilascio 5.11.1 di Amazon EMR](#)
- [Rilascio 5.11.0 di Amazon EMR](#)
- [Rilascio 5.10.1 di Amazon EMR](#)
- [Rilascio 5.10.0 di Amazon EMR](#)
- [Rilascio 5.9.1 di Amazon EMR](#)
- [Rilascio 5.9.0 di Amazon EMR](#)
- [Rilascio 5.8.3 di Amazon EMR](#)
- [Rilascio 5.8.2 di Amazon EMR](#)
- [Rilascio 5.8.1 di Amazon EMR](#)
- [Rilascio 5.8.0 di Amazon EMR](#)
- [Rilascio 5.7.1 di Amazon EMR](#)
- [Rilascio 5.7.0 di Amazon EMR](#)
- [Rilascio 5.6.1 di Amazon EMR](#)

- [Rilascio 5.6.0 di Amazon EMR](#)
- [Rilascio 5.5.4 di Amazon EMR](#)
- [Rilascio 5.5.3 di Amazon EMR](#)
- [Rilascio 5.5.2 di Amazon EMR](#)
- [Rilascio 5.5.1 di Amazon EMR](#)
- [Rilascio 5.5.0 di Amazon EMR](#)
- [Rilascio 5.4.1 di Amazon EMR](#)
- [Rilascio 5.4.0 di Amazon EMR](#)
- [Rilascio 5.3.2 di Amazon EMR](#)
- [Rilascio 5.3.1 di Amazon EMR](#)
- [Rilascio 5.3.0 di Amazon EMR](#)
- [Rilascio 5.2.3 di Amazon EMR](#)
- [Rilascio 5.2.2 di Amazon EMR](#)
- [Rilascio 5.2.1 di Amazon EMR](#)
- [Rilascio 5.2.0 di Amazon EMR](#)
- [Rilascio 5.1.1 di Amazon EMR](#)
- [Rilascio 5.1.0 di Amazon EMR](#)
- [Rilascio 5.0.3 di Amazon EMR](#)
- [Rilascio 5.0.0 di Amazon EMR](#)

Versioni dell'applicazione nei rilasci di Amazon EMR 5.x

Per una tabella completa che elenca le versioni dell'applicazione disponibili in ogni versione di Amazon EMR 5.x, aprire [Versioni delle applicazioni nelle versioni di Amazon EMR 5.x](#) nel browser.

Versione 5.36.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Liviy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
SDK AWS per 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	2.12.15	2.12.15	2.12.10	Non disponibile
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

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.267	0.266	0.261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative alla versione 5.36.1 di Amazon EMR. Le modifiche sono relative alla versione 5.36.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- La versione 5.36.1 di Amazon EMR aggiunge il supporto per l'archiviazione dei log su Amazon S3 durante la riduzione dei cluster. Nelle precedenti versioni 5.x, era possibile archiviare i file di log in Amazon S3 solo durante la terminazione del cluster. Questo miglioramento garantisce la persistenza su Amazon S3 dei file di log generati nel cluster anche dopo la terminazione del nodo. Per ulteriori informazioni, consulta [Configurazione della registrazione e del debug di cluster](#).
- La versione 5.36.1 migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- La versione 5.36.1 riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- La versione 5.36.1 corregge un problema per cui i daemon Amazon EMR sul nodo primario mantenevano i metadati obsoleti per le istanze terminate nel cluster. Il mantenimento di dati obsoleti potrebbe causare un aumento illimitato dell'utilizzo della CPU e della memoria all'interno del cluster e, infine, provocare errori del cluster.
- Per i cluster avviati con più nodi primari, la versione 5.36.1 corregge un problema per cui un guasto hardware di Amazon EC2 su uno dei nodi primari poteva causare l'esito negativo di un secondo nodo primario e rendere instabile il cluster.
- Per i cluster configurati con la crittografia in transito, ora Scalabilità gestita riconosce i dati shuffle di Spark. I dati shuffle di Spark sono dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche. Durante la riduzione, Scalabilità gestita ignora le istanze con dati shuffle. In questo modo si evitano nuovi tentativi di eseguire il processo e nuovi calcoli, che sono costosi in termini di prezzo e prestazioni. Per ulteriori informazioni sulle operazioni di shuffle, consulta la [Guida di programmazione Spark](#).
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	(Centrale), Israele (Tel Aviv) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR

e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.16.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.5.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
<code>emr-s3-dist-cp</code>	2.21.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.7.0	Connettore di EMR S3Select
<code>emrfs</code>	2.51.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.14.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-journalnode	2.10.1-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.9-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.9-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.9-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.9-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.9-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.9-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.9-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.10.1-amzn-1	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-spark	0.10.1-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.10.1-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.13.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	11.0.194	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.267-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.267-amzn-1	Servizio per l'esecuzione di parti di una query.
presto-client	0.267-amzn-1	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.

Componente	Versione	Descrizione
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.8-amzn-2	Client a riga di comando Spark.
spark-history-server	2.4.8-amzn-2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.8-amzn-2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.8-amzn-2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.36.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
hudi-defaults	Modifica i valori nel file hudi-defaults.conf di Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Log delle modifiche

Log delle modifiche per la versione 5.36.1 e note di rilascio

Data	Evento	Descrizione
26/07/2023	Aggiornamento	Nuove etichette della versione del sistema operativo 2.0.20230612.0 e 2.0.20230628.0 .
25/05/2023	Implementazione completata	Amazon EMR 5.36.1 implementato completamente in tutte le regioni supportate

Data	Evento	Descrizione
09/05/2023	Pubblicazione dei documenti	Pubblicazione delle prime note di rilascio di Amazon EMR 5.36.1
04/05/2023	Versione iniziale	Implementazione di Amazon EMR 5.36.1 in alcune regioni commerciali

Rilascio 5.36.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
SDK AWS per 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	2.12.15	2.12.10	Non disponibile	Non disponibile
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	-	-	-	-
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

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.36.0. Le modifiche sono relative alla versione 5.35.0.

Data del rilascio iniziale: 15 giugno 2022

Nuove caratteristiche

- La versione 5.36.0 di Amazon EMR aggiunge il supporto per il linguaggio DDL (Data Definition Language) con Apache Spark su cluster abilitati per Apache Ranger. Ciò consente di utilizzare Apache Ranger per gestire l'accesso a operazioni come la creazione, la modifica e la rimozione di database e tabelle da un cluster Amazon EMR.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR

predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 719.0	4.14.287	10 agosto 2022	Stati Uniti occidentali (California settentrionale), Europa (Parigi), Europa (Stoccolma), Europa (Francoforte), Asia Pacifico (Mumbai), Medio Oriente (Bahrein)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022.0426.0	4.14.281	14 giugno 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Modifiche, miglioramenti e problemi risolti

- Gli aggiornamenti di Amazon EMR 5.36.0 ora supportano: 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, connettore Amazon Glue 1.18.0 e EMRFS 2.51.0.

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. Ca sua della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.16.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.5.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
<code>emr-s3-dist-cp</code>	2.21.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.7.0	Connettore di EMR S3Select

Componente	Versione	Descrizione
emrfs	2.51.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.14.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-4	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.10.1-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.1-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.9-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.9-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.9-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.9-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.9-amzn-2	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.9-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.9-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.10.1-amzn-1	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-spark	0.10.1-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.10.1-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.13.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.
nvidia-cuda	11.0.194	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.267-amzn-1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.267-amzn-1	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
presto-client	0.267-amzn-1	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.8-amzn-2	Client a riga di comando Spark.
spark-history-server	2.4.8-amzn-2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.8-amzn-2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.8-amzn-2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.36.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark	Not available.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Amazon EMR versione 5.35.0

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
SDK AWS per 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	2.12.10	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.14.2	1.13.1	1.12.1	1.12.1

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
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	2.3.9	2.3.8	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.8	2.3.7	2.3.7
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.266	0.261	0.245.1	0.245.1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
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

Note di rilascio

Questa è la nota di rilascio di Amazon EMR 5.35.0.

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.35.0. Le modifiche sono relative alla versione 5.34.0.

Data del rilascio iniziale: 30 marzo 2022

Nuove caratteristiche

- Le applicazioni Amazon EMR versione 5.35 che usano Log4j 1.x e Log4j 2.x sono aggiornate per usare rispettivamente Log4j 1.2.17 (o superiore) e Log4j 2.17.1 (o superiore) e non richiedono l'uso di operazioni di bootstrap per mitigare i problemi CVE nelle versioni precedenti. Per informazioni, consultare [Approccio per mitigare il CVE-2021-44228](#).

Modifiche, miglioramenti e problemi risolti

Modifiche di Flink

Tipo di modifica	Descrizione
Aggiornamenti	<ul style="list-style-type: none"> Aggiorna la versione di flink alla 1.14.2. log4j aggiornato a 2.17.1.

Cambiamenti di Hadoop

Tipo di modifica	Descrizione
Backport open source di Hadoop da EMR 5.34.0	<ul style="list-style-type: none"> • YARN-10438: Gestire null ContainerID in ClientRemService#getContainerReport() • YARN-7266: thread del gestore eventi Timeline Server bloccati • YARN-10438: ATS 1.5 non si avvia se i file RollingLevelDB sono danneggiati o mancanti • HADOOP-13500: Sincronizzazione dell'iter azione dell'oggetto delle proprietà di configurazione • YARN-10651: CapacityScheduler si è bloccato con NPE in AbstractYarnScheduler.updateNodeResource() • HDFS-12221: Sostituisci gli xerces in XmlEditsVisitor • HDFS-16410: analisi Xml non sicura in OfflineEditSXMLLoader
Modifiche e correzioni di Hadoop	<ul style="list-style-type: none"> • Tomcat utilizzato in KMS e HTTPFS è aggiornato a 8.5.75 • In FileSystemOptimizedCommitV2, il marker di successo è stato scritto nel percorso di output CommitJob definito durante la creazione del committer. Poiché i percorsi di output CommitJob e livello di attività possono essere diversi, il percorso è stato corretto per utilizzare quello definito nei file manifest. Per i lavori Hive, ciò comporta che il marker di successo venga scritto correttamente durante l'esecuzione di operazioni come la partizione dinamica o UNION ALL.

Cambiamenti Hive

Tipo di modifica	Descrizione
Hive aggiornato a open source versione 2.3.9 , incluse queste correzioni JIRA	<ul style="list-style-type: none"> • HIVE-17155: findConfFile() in HiveConf.java presenta alcuni problemi con il percorso conf • HIVE-24797: disabilita la convalida dei valori predefiniti durante l'analisi degli schemi Avro • HIVE-21563: Migliora le prestazioni di Table #getEmptyTable disabilita RegisterAllFunctionsOnce • HIVE-18147: i test possono fallire con java.net.bindeXception: indirizzo già in uso • HIVE-24608: torna a get_table nel client HMS per Hive 2.3.x • HIVE-21200: Vectorization - colonna data che lancia java.lang.UnsupportedOperationException per parquet • HIVE-19228: rimuovere l'utilizzo di commons-httpclient 3.x
Hive backport open source da EMR 5.34.0	<ul style="list-style-type: none"> • HIVE-19990: la query con intervallo letterale nella condizione di join non riesce • HIVE-25824: Aggiorna branch-2.3 a log4j 2.17.0 • TEZ-4062: la pianificazione dei tentativi speculativi deve essere interrotta al completamento dell'attività • TEZ-4108: NullPointerException durante la condizione di gara di esecuzione speculativa • TEZ-3918: L'impostazione di tez.task.log.level non funziona
Aggiornamenti e correzioni di Hive	<ul style="list-style-type: none"> • Aggiorna la versione di Log4j a 2.17.1 • Aggiorna la versione di ORC a 1.4.3

Tipo di modifica	Descrizione
	<ul style="list-style-type: none"> Stallo fisso dovuto al thread di penalità in ShuffleScheduler
Nuove funzionalità di	<ul style="list-style-type: none"> Aggiunta la funzione per stampare Hive Query nei registri AM. Questa opzione è disabilitata per impostazione predefinita. Bandiera/Conf:tez.am.emr.print.hive.query.in.log . Stato (predefinito): FALSE.

Modifiche Oozie

Tipo di modifica	Descrizione
Backport open source di Oozie da EMR 5.34.0	<ul style="list-style-type: none"> OOZIE-3652: Oozie launcher dovrebbe riprovare l'elenco delle directory quando si verifica NoSuchFileException

Cambiamenti Pig

Tipo di modifica	Descrizione
Aggiornamenti	<ul style="list-style-type: none"> log4j aggiornato a 1.2.17.

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.

- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere / (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra a e b in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.15.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.5.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.20.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.7.0	Connettore di EMR S3Select
emrfs	2.49.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.14.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.14.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.1-amzn-3	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.10.1-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.10.1-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.9-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.3.9-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.9-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.9-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.9-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.9-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.9-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.9.0-amzn-2	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-spark	0.9.0-amzn-2	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.9.0-amzn-2	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.10.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.13.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.266-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.266-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.266-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.8-amzn-1	Client a riga di comando Spark.
spark-history-server	2.4.8-amzn-1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.8-amzn-1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.8-amzn-1	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.35.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark	Not available.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.34.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
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	2.3.8	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.8	2.3.7	2.3.7	2.3.7
Hudi	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.261	0.245.1	0.245.1	0.240.1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.34.0. Le modifiche sono relative alla versione 5.33.1.

Data del rilascio iniziale: 20 gennaio 2022

Data di rilascio aggiornata: 21 marzo 2022

Nuove caratteristiche

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).
- [Hudi] Miglioramenti per semplificare la configurazione Hudi. Optimistic Concurrency Control disabilitato per impostazione predefinita.

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- In precedenza, il riavvio manuale del gestore delle risorse su un cluster multi-master provocava il ricaricamento, da parte dei daemon su cluster di Amazon EMR come Zookeeper, di tutti i nodi precedentemente disattivati o persi nel file znode di Zookeeper. Ciò causava il superamento dei

limiti predefiniti in determinate situazioni. Ora Amazon EMR rimuove i record dei nodi disattivati o persi più vecchi di un'ora dal file Zookeeper e i limiti interni sono stati aumentati.

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Zeppelin aggiornato alla versione 0.10.0.
- Livy Fix: aggiornato alla versione 0.7.1
- Miglioramento delle prestazioni di Spark: gli esecutori eterogenei sono disabilitati quando alcuni valori di configurazione Spark vengono sovrascritti in EMR 5.34.0.
- I server HttpFS e WebHDFS sono disabilitati per impostazione predefinita. Puoi abilitare nuovamente WebHDFS utilizzando la configurazione di Hadoop, `dfs.webhdfs.enabled`. Il server HttpFS può essere avviato utilizzando `sudo systemctl start hadoop-httpfs`.

Problemi noti

- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione

alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.

- Le query Hue non funzionano in Amazon EMR 6.4.0 perché il server HttpFS di Apache Hadoop è disabilitato per impostazione predefinita. Per utilizzare Hue su Amazon EMR 6.4.0, avvia manualmente il server HttpFS sul nodo primario di Amazon EMR utilizzando `sudo systemctl start hadoop-httpfs` oppure [utilizza una fase di Amazon EMR](#).
- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-`. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni

più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.1	SDK Spark Amazon SageMaker
emr-ddb	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.14.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.4.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.7.0	Connettore di EMR S3Select
emrfs	2.48.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.13.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.13.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-journalnode	2.10.1-amzn-2	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.8-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.8-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.8-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.8-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.8-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.8-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.8-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.9.0-amzn-0	Framework di elaborazione incrementale per alimentare la Data Pipeline a bassa latenza e alta efficienza.
hudi-spark	0.9.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.9.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.4.1	Server multi-utente per notebook Jupyter
livy-server	0.7.1-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.13.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.8.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.1	Client a riga di comando Oozie.
oozie-server	5.2.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.261-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.261-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.261-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.

Componente	Versione	Descrizione
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.8-amzn-0	Client a riga di comando Spark.
spark-history-server	2.4.8-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.8-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.8-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.10.0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.34.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.33.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.33.0/5.33.1. Le modifiche sono relative alla versione 5.32.0.

Data del rilascio iniziale: 19 aprile 2021

Ultimo aggiornamento: 9 agosto 2021

Aggiornamenti

- Connettore Amazon Glue aggiornato alla versione 1.15.0
- AWS SDK for Java aggiornato alla versione 1.11.970
- EMRFS aggiornato alla versione 2.46.0
- EMR Goodies aggiornato alla versione 2.14.0
- EMR Record Server aggiornato alla versione 1.9.0
- EMR S3 Dist CP aggiornato alla versione 2.18.0
- EMR Secret Agent aggiornato alla versione 1.8.0
- Flink aggiornato alla versione 1.12.1
- Hadoop aggiornato alla versione 2.10.1-amzn-1
- Hive aggiornato alla versione 2.3.7-amzn-4
- Hudi aggiornato alla versione 0.7.0
- Hue aggiornato alla versione 4.9.0
- OpenCV aggiornato alla versione 4.5.0
- Presto aggiornato alla versione 0.245.1-amzn-0
- R aggiornato alla versione 4.0.2
- Spark aggiornato alla versione 2.4.7-amzn-1
- TensorFlow aggiornato alla versione 2.4.1
- Zeppelin aggiornato alla versione 0.9.0

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.

- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Configurazione di un cluster per risolvere i problemi di prestazioni di Apache YARN Timeline Server versione 1 e 1.5

Le versioni 1 e 1.5 di Apache YARN Timeline Server possono causare problemi di prestazioni con cluster EMR molto attivi e di grandi dimensioni, in particolare con `yarn.resourcemanager.system-metrics-publisher.enabled=true`, che è l'impostazione predefinita in Amazon EMR. Un YARN Timeline Server v2 open source risolve il problema di prestazioni relativo alla scalabilità di YARN Timeline Server.

Altre soluzioni alternative per questo problema includono:

- La configurazione di `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- L'abilitazione della correzione per questo problema durante la creazione di un cluster, come descritto di seguito.

I seguenti rilasci di Amazon EMR contengono una correzione per questo problema di prestazioni di YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1 e 6.4.x

Per abilitare la correzione su una delle versioni di Amazon EMR sopra specificate, imposta queste proprietà su `true` in un file JSON di configurazioni che viene passato utilizzando il [parametro del comando `aws emr create-cluster`](#): `--configurations file:///./configurations.json`. Oppure abilita la correzione utilizzando l'[interfaccia utente della console di riconfigurazione](#).

Esempio del contenuto del file `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": []
}
]
```

- Adesso, il runtime di Spark è più veloce quando si recuperano le posizioni delle partizioni dal metastore Hive per le query di inserimento Spark.
- Versioni dei componenti aggiornate. Per un elenco delle versioni dei componenti, consulta [Informazioni sui rilasci di Amazon EMR](#) in questa guida.
- Installato il AWS Java SDK Bundle su ogni nuovo cluster. Si tratta di un singolo jar contenente tutti gli SDK di servizio e le relative dipendenze, anziché i singoli jar di componenti. Per ulteriori informazioni, consulta [Dipendenza dell'SDK in bundle](#).
- Risolti i problemi di dimensionamento gestito nelle versioni precedenti di Amazon EMR e apportati miglioramenti in modo da ridurre significativamente le percentuali di errore delle applicazioni.
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Nuove funzionalità di

- Amazon EMR supporta Amazon S3 Access Point, una funzionalità di Amazon S3 che consente di gestire facilmente l'accesso ai data lake condivisi. Utilizzando il tuo alias Amazon S3 Access Point, puoi semplificare l'accesso ai dati su larga scala su Amazon EMR. Puoi utilizzare gli Amazon S3 Access Point con tutte le versioni di Amazon EMR senza alcun costo aggiuntivo in tutte le Regioni AWS in cui è disponibile Amazon EMR. Per ulteriori informazioni sugli alias degli Access Point e degli Amazon S3 Access Point, consulta [Utilizzo di un alias in stile bucket per il punto di accesso](#) nella Guida per l'utente di Amazon S3.
- Amazon EMR-5.33 supporta nuovi tipi di istanza Amazon EC2: c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. Consulta la sezione relativa ai [tipi di istanza supportati](#).

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Per i cluster della sottorete privata di Amazon EMR 6.3.0 e 6.2.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio httpd sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.

• **⚠ Important**

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Il supporto della console per creare una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger non è attualmente abilitato nella Regione GovCloud. La configurazione della sicurezza può essere eseguita utilizzando la CLI. Consulta [Creazione della configurazione di sicurezza EMR](#) nella Guida per la gestione di Amazon EMR.
- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.14.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-notebook-env	1.2.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.6.0	Connettore di EMR S3Select
emrfs	2.46.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.12.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.12.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.10.1-amzn-1.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-1.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-1.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-1.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.10.1-amzn-1.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-1.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-1.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.7-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.7-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-4	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-4	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.7-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.7.0-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.7.0-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.7.0-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.2.2	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MySQL.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.245.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.245.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.245.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.7-amzn-1.1	Client a riga di comando Spark.
spark-history-server	2.4.7-amzn-1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.7-amzn-1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.7-amzn-1.1	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.9.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.33.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark	Not available.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.33.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
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	2.3.7	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

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.14.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.18.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.6.0	Connettore di EMR S3Select
emrfs	2.46.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.12.1	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.12.1	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.1-amzn-1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.10.1-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.3.7-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.7-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-4	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-4	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.7-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.7.0-amzn-1	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.7.0-amzn-1	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.7.0-amzn-1	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.9.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	1.2.2	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.245.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.245.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.245.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	4.0.2	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.7-amzn-1	Client a riga di comando Spark.
spark-history-server	2.4.7-amzn-1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.7-amzn-1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.7-amzn-1	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.4.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.9.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.33.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificazioni	Descrizione	Operazioni di riconfigurazione
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-session	Modifica le impostazioni Flink log4j-session.properties per la sessione Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark	Not available.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modifica le impostazioni di configurazione in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.32.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.0	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
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

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
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

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.

- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I

pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato *myapp-component* con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come *2.2-amzn-2*.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.1	SDK Spark Amazon SageMaker
emr-ddb	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.13.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-notebook-env	1.1.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
emr-s3-dist-cp	2.17.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.6.0	Connettore di EMR S3Select
emrfs	2.45.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.11.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-0.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-0.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-0.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.1-amzn-0.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-0.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-0.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-0.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-0.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.7-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.7-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-3	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-3	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-3	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.7-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.6.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.6.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.6.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.8.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68+	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.240.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.240.1-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.240.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.

Componente	Versione	Descrizione
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.7-amzn-0.1	Client a riga di comando Spark.
spark-history-server	2.4.7-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.7-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.7-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.3.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.32.1

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.
container-executor	Modificare i valori nel file <code>container-executor.cfg</code> di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.
emrfs-site	Modifica le impostazioni EMRFS.	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.

Classificazioni	Descrizione	Operazioni di riconfigurazione
flink-conf	Modifica le impostazioni flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, HiveMetaStore, and MapReduceHistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.	Not available.
presto-env	Modifica i valori nel file presto-env.sh di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file node.properties di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.

Classificazioni	Descrizione	Operazioni di riconfigurazione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark	Not available.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.	Not available.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.32.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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
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

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.32.0. Le modifiche sono relative alla versione 5.31.0.

Data del rilascio iniziale: 8 gennaio 2021

Aggiornamenti

- Connettore Amazon Glue aggiornato alla versione 1.14.0
- Aggiornamento di Amazon SageMaker Spark SDK alla versione 1.4.1
- Aggiornato AWS SDK for Java alla versione 1.11.890

- EMR DynamoDB Connector aggiornato alla versione 4.16.0
- EMRFS aggiornato alla versione 2.45.0
- Parametri di analisi dei log EMR aggiornati alla versione 1.18.0
- Client MetricsAndEventsApiGateway EMR aggiornato alla versione 1.5.0
- EMR Record Server aggiornato alla versione 1.8.0
- EMR S3 Dist CP aggiornato alla versione 2.17.0
- EMR Secret Agent aggiornato alla versione 1.7.0
- Flink aggiornato alla versione 1.11.2
- Hadoop aggiornato alla versione 2.10.1-amzn-0
- Hive aggiornato alla versione 2.3.7-amzn-3
- Hue aggiornato alla versione 4.8.0
- MXNet aggiornato alla versione 1.7.0
- OpenCV aggiornato alla versione 4.4.0
- Presto aggiornato alla versione 0.240.1-amzn-0
- Spark aggiornato alla versione 2.4.7-amzn-0
- TensorFlow aggiornato alla versione 2.3.1

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.

- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Versioni dei componenti aggiornate.
- Per un elenco delle versioni dei componenti, consulta [Informazioni sui rilasci di Amazon EMR](#) in questa guida.

Nuove funzionalità di

- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.
- Stato del supporto IMDS (Instance Metadata Service) V2: Amazon EMR 5.23.1, 5.27.1 e 5.32 o versioni successive utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Per altri rilasci EMR 5.x, la disattivazione di IMDSv1 causa un errore di avvio del cluster.
- A partire da Amazon EMR 5.32.0, è possibile avviare un cluster che si integra nativamente con Apache Ranger. Apache Ranger è un framework open source che consente di abilitare, monitorare e gestire la sicurezza completa dei dati attraverso la piattaforma Hadoop. Per ulteriori informazioni, consulta [Apache Ranger](#). L'integrazione nativa consente di utilizzare Apache Ranger per imporre un controllo granulare di accesso ai dati su Amazon EMR. Consulta [Integrazione di Amazon EMR con Apache Ranger](#) nella Guida ai rilasci di Amazon EMR.
- Amazon EMR 5.32.0 supporta Amazon EMR su EKS. Per ulteriori dettagli su come iniziare a utilizzare EMR su EKS, consulta [Che cos'è Amazon EMR su EKS?](#).
- Amazon EMR 5.32.0 supporta Amazon EMR Studio (anteprima). Per ulteriori informazioni sulle nozioni di base di EMR Studio, consulta [Amazon EMR Studio \(anteprima\)](#).

- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).

Problemi noti

- Per i cluster della sottorete privata di Amazon EMR 6.3.0 e 6.2.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio httpd sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.
- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le

istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Il supporto della console per creare una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger non è attualmente abilitato nella Regione GovCloud. La configurazione della sicurezza può essere eseguita utilizzando la CLI. Consulta [Creazione della configurazione di sicurezza EMR](#) nella Guida per la gestione di Amazon EMR.
- Quando la crittografia AtRestEncryption o HDFS è abilitata in un cluster che utilizza Amazon EMR 5.31.0 o 5.32.0, le query Hive determinano la seguente eccezione di runtime.

```
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
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.16.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.13.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.1.0	Ambiente Conda per EMR Notebooks che include il gateway aziendale Jupyter
<code>emr-s3-dist-cp</code>	2.17.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.6.0	Connettore di EMR S3Select

Componente	Versione	Descrizione
emrfs	2.45.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.2	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.2	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.10.1-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.1-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.1-amzn-0	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.10.1-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.1-amzn-0	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.10.1-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.1-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.1-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.1-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.1-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.7-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.7-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-3	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-3	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.7-amzn-3	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.7-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.6.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.6.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.6.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.8.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.7.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.68	Server di database MySQL.
nvidia-cuda	10.1.243	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.240.1-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.240.1-amzn-0	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
presto-client	0.240.1-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.7-amzn-0	Client a riga di comando Spark.
spark-history-server	2.4.7-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.7-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.7-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.3.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Le azioni di riconfigurazione vengono eseguite quando si specifica una configurazione per gruppi di istanze in un cluster in esecuzione. Amazon EMR avvia solo le azioni di riconfigurazione per le classificazioni modificate. Per ulteriori informazioni, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Classificazioni emr-5.32.0

Classificazioni	Descrizione	Operazioni di riconfigurazione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.	Restarts the ResourceManager service.

Classificazioni	Descrizione	Operazioni di riconfigurazione
container-executor	Modificare i valori nel file container-executor.cfg di Hadoop YARN.	Not available.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.	Not available.
core-site	Modifica i valori nel file core-site.xml di 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	Modifica le impostazioni relative a docker.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
emrfs-site	Modifica le impostazioni EMRFS.	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	Modifica le impostazione flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifica le impostazioni Flink log4j.properties.	Restarts Flink history server.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.	Restarts Flink history server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifica i valori nel file log4j.properties di 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	Modifica la configurazione server ssl hadoop	Not available.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hbase	Impostazioni Amazon EMR per Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Modifica i valori nell'ambiente HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.	Not available.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.	Should not be reconfigured.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.	Not available.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.	Not available.
hive-env	Modifica i valori nell'ambiente Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.	Restarts HiveServer2 and HiveMetastore.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.	Not available.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.	Not available.
hive-site	Modifica i valori nel file hive-site.xml di Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2	Not available.
hue-ini	Modifica i valori nel file ini di Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.	Restarts Hadoop Httpfs service.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.	Not available.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifica i valori nell'ambiente Hudi.	Not available.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.	Not available.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.	Not available.
livy-conf	Modifica i valori nel file livy.conf di Livy.	Restarts Livy Server.
livy-env	Modifica i valori nell'ambiente Livy.	Restarts Livy Server.
livy-log4j	Modifica le impostazioni Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classificazioni	Descrizione	Operazioni di riconfigurazione
oozie-env	Modifica i valori nell'ambiente di Oozie.	Restarts Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.	Restarts Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.	Not available.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.	Not available.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.	Not available.
pig-env	Change values in the Pig environment.	Not available.
pig-properties	Modifica i valori nel file pig.properties di Pig.	Restarts Oozie.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.	Not available.
presto-log	Modifica i valori nel file log.properties di Presto.	Restarts Presto-Server.
presto-config	Modifica i valori nel file config.properties di Presto.	Restarts Presto-Server.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.	Not available.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.	Restarts Presto-Server.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.	Not available.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.	Not available.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.	Not available.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.	Not available.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.	Not available.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.	Not available.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.	Not available.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.	Not available.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.	Not available.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.	Not available.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.	Not available.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.	Not available.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.	Not available.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.	Not available.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.	Not available.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.	Restarts EMR record server.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.	Restarts EMR record server.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.	Restarts EMR record server.
spark	Impostazioni Amazon EMR per Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifica i valori nell'ambiente Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark	Not available.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.	Not available.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.	Not available.

Classificazioni	Descrizione	Operazioni di riconfigurazione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.	Not available.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.	Restarts Oozie.
yarn-env	Modifica i valori nell'ambiente YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.	Restarts Zookeeper server.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.	Restarts Zookeeper server.

Rilascio 5.31.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)

- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Hive	2.3.7	2.3.7	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.7.1	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.238.3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).

- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-`. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato *myapp-component* con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.4.0	SDK Spark Amazon SageMaker
emr-ddb	4.15.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.13.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.15.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.6.0	Connettore di EMR S3Select
emrfs	2.43.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.

Componente	Versione	Descrizione
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.10.0-amzn-0.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.0-amzn-0.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.0-amzn-0.1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.0-amzn-0.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.10.0-amzn-0.1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.10.0-amzn-0.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.0-amzn-0.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.7-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.7-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.7-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hudi	0.6.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.6.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.6.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.7.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.6.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.238.3-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.238.3-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.238.3-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	2.4.6-amzn-0.1	Client a riga di comando Spark.
spark-history-server	2.4.6-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.6-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.6-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.1.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.31.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.31.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Hudi	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.6.0	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.31.0. Le modifiche sono relative alla versione 5.30.1.

Data del rilascio iniziale: 9 ottobre 2020

Ultimo aggiornamento: 15 ottobre 2020

Aggiornamenti

- Connettore Amazon Glue aggiornato alla versione 1.13.0
- Amazon SageMaker Spark SDK aggiornato alla versione 1.4.0
- Connettore Amazon Kinesis aggiornato alla versione 3.5.9
- Aggiornato AWS SDK for Java alla versione 1.11.852
- Bigtop-tomcat aggiornato alla versione 8.5.56
- EMR FS aggiornato alla versione 2.43.0
- Client MetricsAndEventsApiGateway EMR aggiornato alla versione 1.4.0
- EMR S3 Dist CP aggiornato alla versione 2.15.0
- EMR S3 Select aggiornato alla versione 1.6.0
- Flink aggiornato alla versione 1.11.0
- Hadoop aggiornato alla versione 2.10.0
- Hive aggiornato alla versione 2.3.7
- Hudi aggiornato alla versione 0.6.0
- Hue aggiornato alla versione 4.7.1
- Aggiornato JupyterHub alla versione 1.1.0
- MXNet aggiornato alla versione 1.6.0
- OpenCV aggiornato alla versione 4.3.0
- Presto aggiornato alla versione 0.238.3
- TensorFlow aggiornato alla versione 2.1.0

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Le [statistiche di colonna Hive](#) sono supportate per Amazon EMR 5.31.0 e versioni successive.
- Versioni dei componenti aggiornate.
- Supporto di EMRFS S3EC V2 in Amazon EMR 5.31.0. In S3 Java SDK 1.11.837 e versioni successive, il client di crittografia versione 2 (S3EC V2) è stato introdotto con vari miglioramenti alla sicurezza. Per ulteriori informazioni, consulta gli argomenti seguenti:
 - Post del blog S3: [Updates to the Amazon S3 encryption client \(Aggiornamenti al client di crittografia Amazon S3\)](#).
 - AWS SDK for Java Guida per gli sviluppatori: [Migrazione dei client di crittografia e decrittografia a V2](#).
 - Guida alla gestione di EMR: [Crittografia lato client di Amazon S3](#).

Il client di crittografia V1 è ancora disponibile nell'SDK per la compatibilità con le versioni precedenti.

Nuove funzionalità di

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```


Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Con Amazon EMR 5.31.0, puoi avviare un cluster che si integra con Lake Formation. L'integrazione fornisce un filtraggio dei dati granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Inoltre, consente l'accesso federato single sign-on a EMR Notebooks o Apache Zeppelin da un sistema di identità aziendale. Per ulteriori informazioni, consulta [Integrazione di Amazon EMR con AWS Lake Formation](#) nella Guida alla gestione di Amazon EMR.

Amazon EMR con Lake Formation è attualmente disponibile in 16 regioni AWS: Stati Uniti orientali (Ohio e N. Virginia), Stati Uniti occidentali (California settentrionale e Oregon), Asia Pacifico (Mumbai, Seoul, Singapore, Sydney e Tokyo), Canada (Centrale), Europa (Francoforte, Irlanda, Londra, Parigi e Stoccolma), Sud America (San Paolo).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Quando la crittografia AtRestEncryption o HDFS è abilitata in un cluster che utilizza Amazon EMR 5.31.0 o 5.32.0, le query Hive determinano la seguente eccezione di runtime.

```
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
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:

- Due o più partizioni vengono scansionate dalla stessa tabella.
- Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-`. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.4.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.15.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.13.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.15.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.6.0	Connettore di EMR S3Select
emrfs	2.43.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.11.0	Applicazioni e script client a riga di comando Apache Flink.
flink-jobmanager-config	1.11.0	Gestione delle risorse sui nodi EMR per Apache Flink JobManager.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.10.0-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.10.0-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.10.0-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.10.0-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.10.0-amzn-0	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.10.0-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.10.0-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.10.0-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.10.0-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.10.0-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.10.0-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.7-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.7-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.7-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.7-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.7-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.7-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.6.0-amzn-0	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-spark	0.6.0-amzn-0	Libreria bundle per eseguire Spark con Hudi.
hudi-presto	0.6.0-amzn-0	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.7.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter

Componente	Versione	Descrizione
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.6.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.238.3-amzn-0	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.238.3-amzn-0	Servizio per l'esecuzione di parti di una query.
presto-client	0.238.3-amzn-0	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.6-amzn-0	Client a riga di comando Spark.
spark-history-server	2.4.6-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.6-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.6-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	2.1.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.31.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop.
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.

Classificazioni	Descrizione
hbase-metrics	Modificare valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.

Classificazioni	Descrizione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.

Classificazioni	Descrizione
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.30.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
SDK AWS per Java	1.11.759	1.11.759	1.11.759	1.11.682

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
Python	2.7, 3.7	2.7, 3.7	2.7, 3.7	2.7, 3.6
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.10.0	1.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.6.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
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

Note di rilascio

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

Modifiche, miglioramenti e problemi risolti

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.

- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-`. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.3.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.13.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emr-s3-select	1.5.0	Connettore di EMR S3Select
emrfs	2.40.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.10.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-6.1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-6.1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-6.1	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.5-amzn-6.1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.8.5-amzn-6.1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-6.1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-6.1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.6-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.6-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.6-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.6-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.2-incubating	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.2-incubating	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.6.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.

Componente	Versione	Descrizione
mariadb-server	5.5.64+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.232	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.232	Servizio per l'esecuzione di parti di una query.
presto-client	0.232	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.

Componente	Versione	Descrizione
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.5-amzn-0.1	Client a riga di comando Spark.
spark-history-server	2.4.5-amzn-0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.5-amzn-0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.5-amzn-0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.30.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.

Classificazioni	Descrizione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.

Classificazioni	Descrizione
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.30.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.9.1	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.30.1. Le modifiche sono relative alla versione 5.30.0.

Data del rilascio iniziale: 30 giugno 2020

Ultimo aggiornamento: 24 agosto 2020

Modifiche, miglioramenti e problemi risolti

- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Risolto il problema per cui il processo del controller di istanza generava un numero infinito di processi.
- Risolto un problema per cui Hue non era in grado di eseguire una query Hive, mostrando il messaggio "database is locked (database bloccato)" e impedendo l'esecuzione di query.
- Risolto un problema Spark per consentire l'esecuzione simultanea di più attività nel cluster EMR.
- Risolto un problema del notebook Jupyter che causava l'errore "too many open files (troppi file aperti)" nel server Jupyter.
- Corretto un problema relativo all'ora di inizio del cluster.

Nuove funzionalità di

- Le interfacce utente delle applicazioni persistenti Timeline Server di YARN e Tez sono disponibili con Amazon EMR versioni 6.x ed EMR versione 5.30.1 e successive. L'accesso ai collegamenti One-Click alla cronologia dell'applicazione persistente consente di accedere rapidamente alla cronologia dei processi senza configurare un proxy Web tramite una connessione SSH. I log dei cluster attivi e terminati sono disponibili per 30 giorni dopo la fine dell'applicazione. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce utente delle applicazioni persistenti](#) nella Guida alla gestione di Amazon EMR.
- Le API di esecuzione di EMR Notebooks sono disponibili per eseguire i notebook EMR tramite uno script o una riga di comando. La possibilità di avviare, arrestare, elencare e descrivere le esecuzioni dei notebook EMR senza la console AWS consente di controllare a livello di codice un notebook EMR. Utilizzando una cella del notebook parametrizzata, è possibile passare diversi

valori di parametro a un notebook senza doverne creare una copia per ogni nuovo set di valori dei parametri. Consulta [Operazioni dell'API EMR](#). Per un codice di esempio, consulta [Comandi di esempio per eseguire EMR Notebooks a livello di codice](#).

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```


Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- EMR Notebooks

La caratteristica che consente di installare kernel e librerie Python aggiuntive sul nodo primario del cluster è disabilitata per impostazione predefinita in EMR versione 5.30.1. Per ulteriori informazioni su questa caratteristica, consulta la sezione [Installing Kernels and Python Libraries on a Cluster Primary Node](#) (Installazione di kernel e librerie Python su un nodo primario del cluster).

Per abilitare questa funzione, procedi come segue:

1. Assicurati che le policy di autorizzazione associate al ruolo di servizio per EMR Notebooks consentano l'operazione seguente:

```
elasticmapreduce:ListSteps
```

Per ulteriori informazioni, consultare il [Ruolo del servizio per EMR Notebooks](#).

- Utilizza la AWS CLI per eseguire un passaggio sul cluster che imposta EMR Notebooks, come mostrato nell'esempio seguente. Sostituisci *us-east-1* con la Regione in cui risiede il cluster. Per ulteriori informazioni, consulta [Aggiunta di fasi a un cluster utilizzando la AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- Dimensionamento gestito

Le operazioni di dimensionamento gestito su cluster 5.30.0 e 5.30.1 senza Presto installato possono causare errori delle applicazioni o far sì che un gruppo di istanze o un parco istanze uniforme mantenga lo stato ARRESTED, in particolare quando un'operazione di dimensionamento verso il basso è seguita rapidamente da un'operazione di dimensionamento verso l'alto.

Come soluzione alternativa, scegli Presto come applicazione da installare quando crei un cluster con Amazon EMR rilasci 5.30.0 e 5.30.1, anche se il tuo processo non richiede Presto.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.3.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.13.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.5.0	Connettore di EMR S3Select
<code>emrfs</code>	2.40.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.10.0	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-6	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.8.5-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.5-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.6-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.6-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.6-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.2-incubating	Framework di elaborazione incrementale per alimentar e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.2-incubating	Libreria bundle per eseguire Presto con Hudi.

Componente	Versione	Descrizione
hue-server	4.6.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.232	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.232	Servizio per l'esecuzione di parti di una query.
presto-client	0.232	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger
spark-client	2.4.5-amzn-0	Client a riga di comando Spark.
spark-history-server	2.4.5-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.4.5-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.5-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.30.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.

Classificazioni	Descrizione
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file <code>jupyterhub_config.py</code> di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file <code>config.json</code> di Sparkmagic.
livy-conf	Modifica i valori nel file <code>livy.conf</code> di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy <code>log4j.properties</code> .
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file <code>mapred-site.xml</code> dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file <code>oozie-log4j.properties</code> di Oozie.
oozie-site	Modifica i valori nel file <code>oozie-site.xml</code> di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.

Classificazioni	Descrizione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.

Classificazioni	Descrizione
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.30.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.10.0	1.9.1	1.9.0	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.0.0	1.0.0	1.0.0
Livy	0.7.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.30.0. Le modifiche sono relative alla versione 5.29.0.

Data del rilascio iniziale: 13 marzo 2020

Ultimo aggiornamento: 25 giugno 2020

Aggiornamenti

- Aggiornato AWS SDK for Java alla versione 1.11.759
- Aggiornamento di Amazon SageMaker Spark SDK alla versione 1.3.0
- Aggiornato EMR Record Server alla versione 1.6.0
- Flink aggiornato alla versione 1.10.0

- Aggiornato Ganglia alla versione 3.7.2
- HBase aggiornato alla versione 1.4.13
- Aggiornato Hudi alla versione 0.5.2-incubating
- Hue aggiornato alla versione 4.6.0
- Aggiornato JupyterHub alla versione 1.1.0
- Livy aggiornato alla versione 0.7.0-incubating
- Aggiornato Oozie alla versione 5.2.0
- Presto aggiornato alla versione 0.232
- Aggiornato Spark alla versione 2.4.5
- Connettori e driver aggiornati: connettore Amazon Glue 1.12.0; connettore Amazon Kinesis 3.5.0; connettore EMR DynamoDB 4.14.0

Nuove funzionalità di

- EMR Notebooks: se utilizzato con cluster EMR creati con 5.30.0, i kernel di EMR Notebooks vengono eseguiti sul cluster. Ciò migliora le prestazioni dei notebook e consente di installare e personalizzare i kernel. Puoi anche installare librerie Python sul nodo primario del cluster. Per ulteriori informazioni, consulta l'argomento relativo all'[installazione e l'utilizzo di kernel e librerie](#) nella Guida alla gestione di EMR.
- Dimensionamento gestito: con la versione 5.30.0 e successive di Amazon EMR, puoi abilitare il dimensionamento gestito da EMR per aumentare o diminuire automaticamente il numero di istanze o unità nel cluster in base al carico di lavoro. Amazon EMR valuta continuamente i parametri dei cluster per prendere decisioni di dimensionamento che ottimizzano i cluster in termini di costi e velocità. Per maggiori informazioni, consulta [Dimensionamento delle risorse del cluster](#) nella Guida alla gestione di Amazon EMR.
- Crittografia dei file di log archiviati in Amazon S3: con Amazon EMR versione 5.30.0 e successive, puoi crittografare i file di log archiviati in Amazon S3 con una chiave gestita dal cliente AWS KMS. Per ulteriori informazioni, consulta [Crittografia dei file di log archiviati in Amazon S3](#) nella Guida alla gestione di Amazon EMR.
- Supporto di Amazon Linux 2: nella versione EMR 5.30.0 e successive, EMR utilizza il sistema operativo Amazon Linux 2. Le nuove AMI personalizzate (Amazon Machine Image) devono essere basate sull'AMI Amazon Linux 2. Per ulteriori informazioni, consulta [Utilizzo di un AMI personalizzato](#).

- Scalabilità automatica Presto con tolleranza: i cluster EMR che usano 5.30.0 possono essere impostati con un periodo di timeout di scalabilità automatica che consente alle attività Presto di terminare l'esecuzione prima che il loro nodo venga disattivato. Per ulteriori informazioni, consulta [Utilizzo della scalabilità automatica Presto con disattivazione con tolleranza](#).
- Creazione di parchi istanze con nuova opzione di strategia di allocazione: una nuova opzione di strategia di allocazione è disponibile in EMR versione 5.12.1 e successive. Offre un provisioning dei cluster più rapido, un'allocazione di istanze Spot più accurata e una minore interruzione delle istanze Spot. Sono necessari aggiornamenti ai ruoli di servizio EMR non predefiniti. Consulta [Configurazione di parchi istanze](#).
- Comandi `sudo systemctl stop` e `sudo systemctl start`: nella versione EMR 5.30.0 e successive, le quali usano il sistema operativo Amazon Linux 2 OS, EMR utilizza i comandi `sudo systemctl stop` e `sudo systemctl start` per riavviare i servizi. Per ulteriori informazioni, consulta [Come riavviare un servizio in Amazon EMR?](#).

Modifiche, miglioramenti e problemi risolti

- EMR versione 5.30.0 non installa Ganglia per impostazione predefinita. È possibile selezionare esplicitamente Ganglia da installare quando si crea un cluster.
- Ottimizzazione delle prestazioni di Spark.
- Ottimizzazione delle prestazioni di Presto.
- Python 3 è l'impostazione predefinita per Amazon EMR 5.30.0 e versioni successive.
- Il gruppo di sicurezza gestito predefinito per l'accesso al servizio nelle sottoreti private è stato aggiornato con nuove regole. Se si utilizza un gruppo di sicurezza personalizzato per l'accesso al servizio, è necessario includere le stesse regole del gruppo di sicurezza gestito predefinito. Per ulteriori informazioni, consulta [Gruppo di sicurezza gestito da Amazon EMR per l'accesso al servizio \(sottoreti private\)](#). Se si utilizza un ruolo di servizio personalizzato per Amazon EMR, è necessario concedere l'autorizzazione a `ec2:describeSecurityGroups` per consentire a EMR di convalidare se i gruppi di sicurezza sono stati creati correttamente. Se si utilizza `EMR_DefaultRole`, questa autorizzazione è già inclusa nella policy gestita di default.

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` ed `emr-6.2.0` si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno

un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
```

```
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Dimensionamento gestito

Le operazioni di dimensionamento gestito su cluster 5.30.0 e 5.30.1 senza Presto installato possono causare errori delle applicazioni o far sì che un gruppo di istanze o un parco istanze uniforme mantenga lo stato ARRESTED, in particolare quando un'operazione di dimensionamento verso il basso è seguita rapidamente da un'operazione di dimensionamento verso l'alto.

Come soluzione alternativa, scegli Presto come applicazione da installare quando crei un cluster con Amazon EMR rilasci 5.30.0 e 5.30.1, anche se il tuo processo non richiede Presto.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```


In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Il motore di database predefinito per Hue 4.6.0 è SQLite, che causa problemi quando si tenta di utilizzare Hue con un database esterno. Per risolvere questo problema, imposta `engine` nella tua classificazione di configurazione `hue-ini` su `mysql`. Questo problema è stato risolto nella versione 5.30.1 di Amazon EMR.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR

e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.3.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.14.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.13.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.5.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-notebook-env</code>	1.0.0	Ambiente Conda per Notebook EMR
<code>emr-s3-dist-cp</code>	2.14.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.5.0	Connettore di EMR S3Select
<code>emrfs</code>	2.40.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.10.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-6	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.

Componente	Versione	Descrizione
hadoop-https-server	2.8.5-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.13	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.13	Servizio per servire una o più regioni HBase.
hbase-client	1.4.13	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.13	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.13	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.6-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.6-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.6-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hudi	0.5.2-incubating	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.2-incubating	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.6.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.1.0	Server multi-utente per notebook Jupyter
livy-server	0.7.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mariadb-server	5.5.64	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.2.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.232	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.232	Servizio per l'esecuzione di parti di una query.
presto-client	0.232	Presto client della riga di comando che viene installato su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.3	The R Project for Statistical Computing
ranger-kms-server	1.2.0	Sistema di gestione delle chiavi Apache Ranger

Componente	Versione	Descrizione
spark-client	2.4.5-amzn-0	Client a riga di comando Spark.
spark-history-server	2.4.5-amzn-0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.5-amzn-0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.5-amzn-0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.30.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
hudi-env	Modifica i valori nell'ambiente Hudi.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.29.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.9.1	1.9.0	1.9.0	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.5

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
Hudi	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.227	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.29.0. Le modifiche sono relative alla versione 5.28.1.

Data del rilascio iniziale: 17 gennaio 2020

Aggiornamenti

- Aggiornato AWS SDK for Java alla versione 1.11.682
- Hive aggiornato alla versione 2.3.6
- Flink aggiornato alla versione 1.9.1
- EmrFS aggiornato alla versione 2.38.0
- EMR DynamoDB Connector aggiornato alla versione 4.13.0

Modifiche, miglioramenti e problemi risolti

- Spark
 - Ottimizzazione delle prestazioni di Spark.
- EMRFS
 - La Guida alla gestione aggiorna le impostazioni predefinite di `emrfs-site.xml` per una visualizzazione coerente.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.

- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.6	SDK Spark Amazon SageMaker

Componente	Versione	Descrizione
emr-ddb	4.13.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.12.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.13.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.4.0	Connettore di EMR S3Select
emrfs	2.38.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.9.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.8.5-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-5	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.8.5-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.5-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.6-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.6-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.6-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.0-incubating	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.0-incubating	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.227	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.227	Servizio per l'esecuzione di parti di una query.
presto-client	0.227	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.29.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.28.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.5	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.224	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.28.1. Le modifiche sono relative alla versione 5.28.0.

Data del rilascio iniziale: 10 gennaio 2020

Modifiche, miglioramenti e problemi risolti

- Spark
 - Risolti i problemi di compatibilità di Spark.
- Parametri di CloudWatch
 - È stata corretta la pubblicazione dei parametri di Amazon CloudWatch su un cluster EMR con più nodi primari.
- Messaggio di log disattivato
 - Messaggio di falso log disabilitato, "[...] utilizzando la vecchia versione (<4.5.8) del client http Apache".

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.6	SDK Spark Amazon SageMaker

Componente	Versione	Descrizione
emr-ddb	4.12.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.11.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.13.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.37.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.9.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.8.5-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-5	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.8.5-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.5-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.6-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.6-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.6-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.0-incubating	Framework di elaborazione incrementale per alimentare e la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.0-incubating	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.227	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.227	Servizio per l'esecuzione di parti di una query.
presto-client	0.227	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.28.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.28.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.5	2.3.5	2.3.5
Hudi	0.5.0-incubating	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.2	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
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)	-	-	-	-

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.28.0. Le modifiche sono relative alla versione 5.27.0.

Data del rilascio iniziale: 12 novembre 2019

Aggiornamenti

- Flink aggiornato alla versione 1.9.0
- Hive aggiornato alla versione 2.3.6
- MXNet aggiornato alla versione 1.5.1
- Phoenix aggiornato alla versione 4.14.3
- Presto aggiornato alla versione 0.227
- Zeppelin aggiornato alla versione 0.8.2

Nuove funzionalità di

- [Apache Hudi](#) è ora disponibile per Amazon EMR per l'installazione durante la creazione di un cluster. Per ulteriori informazioni, consulta [Hudi](#).
- (25 novembre 2019) È ora possibile scegliere di eseguire più fasi in parallelo per migliorare l'utilizzo dei cluster e risparmiare sui costi. È anche possibile annullare sia le fasi in attesa che quelle in esecuzione. Per ulteriori informazioni, consultare [Utilizzo di fasi mediante l'AWS CLI e la console](#).
- (3 dicembre 2019) Adesso è possibile creare ed eseguire cluster EMR su AWS Outposts. AWS Outposts abilita i servizi nativi di, l'infrastruttura e i modelli operativi di AWS nelle strutture in locale. Negli ambienti AWS Outposts è possibile utilizzare le stesse API, strumenti e infrastruttura AWS utilizzati in AWS Cloud. Per ulteriori informazioni, consulta [Cluster EMR su AWS Outposts](#).
- (11 marzo 2020) A partire dalla versione Amazon EMR 5.28.0, è possibile creare ed eseguire i cluster Amazon EMR in una sottorete di AWS Local Zones come estensione logica di una Regione AWS che supporta le Local Zones. Una Local Zone consente ad Amazon EMR di individuare

le funzionalità e un subset di servizi AWS, come i servizi di calcolo e archiviazione, per essere il più possibile vicina agli utenti, fornendo accesso a latenza molto bassa alle applicazioni in esecuzione localmente. Per un elenco delle Local Zones disponibili, consulta [AWS Local Zones](#). Per informazioni sull'accesso disponibile per le Local Zones AWS, consulta [Regioni, zone di disponibilità e Local Zones](#).

Attualmente, Local Zones non supporta Amazon EMR Notebooks e non supporta connessioni effettuate direttamente ad Amazon EMR utilizzando l'interfaccia endpoint VPC (AWS PrivateLink).

Modifiche, miglioramenti e problemi risolti

- Ampliato il supporto delle applicazioni per i cluster a elevata disponibilità
 - Per ulteriori informazioni, consulta la sezione [Applicazioni supportate in un cluster EMR con più nodi primari](#) nella Guida alla gestione di Amazon EMR.
- Spark
 - Ottimizzazione delle prestazioni
- Hive
 - Ottimizzazione delle prestazioni
- Presto
 - Ottimizzazione delle prestazioni

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.


```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.6	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.12.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.11.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.13.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.37.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.9.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.5-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-5	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.8.5-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.8.5-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.6-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.6-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.6-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.6-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.6-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.6-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hudi	0.5.0-incubating	Framework di elaborazione incrementale per alimentare la pipeline dei dati a bassa latenza e alta efficienza.
hudi-presto	0.5.0-incubating	Libreria bundle per eseguire Presto con Hudi.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.5.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.3-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.227	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.227	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
presto-client	0.227	Presto client della riga di comando che viene installato o su master di standby di un cluster HA dove server Presto non viene avviato.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.28.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.27.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Liv](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
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

Note di rilascio

Questo è un rilascio di patch. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Stato del supporto IMDS (Instance Metadata Service) V2: Amazon EMR 5.23.1, 5.27.1 e 5.32 o versioni successive utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Per altri rilasci EMR 5.x, la disattivazione di IMDSv1 causa un errore di avvio del cluster.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.4	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.12.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.11.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.13.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.36.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.8.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-https-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.5-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.5-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.5-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.5-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.5-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.5-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.2-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.224	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.224	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.

Componente	Versione	Descrizione
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.27.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.

Classificazioni	Descrizione
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.

Classificazioni	Descrizione
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.

Classificazioni	Descrizione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.

Classificazioni	Descrizione
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.27.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.8.1	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.27.0. Le modifiche sono relative alla versione 5.26.0.

Data del rilascio iniziale: 23 settembre 2019

Aggiornamenti

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0

- Spark 2.4.4
- TensorFlow 1.14.0
- Connettori e driver:
 - DynamoDB Connector 4.12.0

Nuove funzionalità di

- (24 ottobre 2019) Le seguenti nuove caratteristiche in EMR Notebooks sono disponibili con tutte le versioni di Amazon EMR
 - Ora è possibile associare repository Git ai notebook EMR per archiviare i notebook in un ambiente controllato dalla versione. È possibile condividere il codice con i peer e riutilizzare i notebook Jupyter esistenti tramite repository Git remoti. Per ulteriori informazioni, consulta [Associazione di repository Git ad Amazon EMR Notebooks](#) nella Guida alla gestione di Amazon EMR.
 - L'[utility nbtime](#) è ora disponibile in EMR Notebooks per semplificare il confronto e la fusione di notebook.
 - Adesso, EMR Notebooks supporta JupyterLab. JupyterLab è un ambiente di sviluppo interattivo basato sul Web completamente compatibile con i notebook Jupyter. Ora è possibile scegliere di aprire il notebook nell'editor del notebook Jupyter o JupyterLab.
- (30 ottobre 2019) Con Amazon EMR versione 5.25.0 e successive, è possibile connettersi all'interfaccia utente di Spark History Server dalla pagina Summary (Riepilogo) del cluster o dalla scheda Application history (Cronologia applicazioni) nella console. Anziché configurare un proxy Web tramite una connessione SSH, puoi accedere rapidamente all'interfaccia utente del server della cronologia Spark per visualizzare i parametri dell'applicazione e accedere ai file di log pertinenti per cluster attivi e terminati. Per ulteriori informazioni, consulta [Accesso fuori cluster alle interfacce utente dell'applicazione persistente](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Cluster Amazon EMR con più nodi primari
 - È possibile installare ed eseguire Flink su un cluster Amazon EMR con più nodi primari. Per ulteriori informazioni, consulta [Applicazioni e funzionalità supportate](#).
 - È possibile configurare la crittografia trasparente HDFS su un cluster Amazon EMR con più nodi primari. Per maggiori informazioni, consulta la sezione [Crittografia trasparente HDFS su cluster EMR con più nodi primari](#).

- È ora possibile modificare la configurazione delle applicazioni in esecuzione su un cluster Amazon EMR con più nodi primari. Per ulteriori informazioni, consulta [Specifiche di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).
- Connettore Amazon EMR-DynamoDB
 - Adesso, il connettore Amazon EMR-DynamoDB supporta i seguenti tipi di dati DynamoDB: booleano, elenco, mappa, elemento, null. Per ulteriori informazioni, consulta [Impostazione di una tabella Hive per l'esecuzione dei comandi Hive](#).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.4	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.12.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.11.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.13.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.3.0	Connettore di EMR S3Select
<code>emrfs</code>	2.36.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.8.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.

Componente	Versione	Descrizione
hadoop-https-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.5-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.5-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.5-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.5-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.5-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.5-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	1.0.0	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.2-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.224	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.224	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.4	Client a riga di comando Spark.
spark-history-server	2.4.4	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.4	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.4	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.14.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.27.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.
ranger-kms-dbks-site	Modifica i valori nel file <code>dbks-site.xml</code> di Ranger KMS.
ranger-kms-site	Modifica i valori nel file <code>ranger-kms-site.xml</code> di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file <code>server.properties</code> di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file <code>log4j.properties</code> di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.26.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
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
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)	-	-	-	-

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.26.0. Le modifiche sono relative alla versione 5.25.0.

Data del rilascio iniziale: 8 agosto 2019

Ultimo aggiornamento: 19 agosto 2019

Aggiornamenti

- AWS SDK for Java 1.11.595
- HBase 1.4.10
- Phoenix 4.14.2
- Connettori e driver:
 - DynamoDB Connector 4.11.0
 - MariaDB Connector 2.4.2
 - Driver JDBC Amazon Redshift 1.2.32.1056

Nuove funzionalità di

- (Beta) Con Amazon EMR 5.26.0 puoi avviare un cluster che si integra con Lake Formation. L'integrazione fornisce un accesso granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Inoltre, consente l'accesso federato single sign-on a EMR Notebooks o Apache Zeppelin da un sistema di identità aziendale. Per ulteriori informazioni, consulta [Integrazione di Amazon EMR con AWS Lake Formation \(Beta\)](#).
- (19 agosto 2019) Il blocco dell'accesso pubblico in Amazon EMR è ora disponibile con tutti i rilasci di Amazon EMR che supportano i gruppi di sicurezza. Il blocco dell'accesso pubblico è un'impostazione a livello di account applicata a ogni Regione AWS. Il blocco degli accessi pubblici impedisce l'avvio di un cluster quando per un gruppo di sicurezza a esso associato è impostata una regola che consente il traffico in entrata da IPv4 0.0.0.0/0 o IPv6 ::/0 (accesso pubblico) su

una porta, a meno che la porta non sia specificata come eccezione. La porta 22 è un'eccezione per impostazione predefinita. Per ulteriori informazioni, consulta [Utilizzo del blocco dell'accesso pubblico in Amazon EMR](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Notebook EMR
 - Con EMR 5.26.0 e versioni successive, EMR Notebooks supporta librerie Python con ambito notebook oltre alle librerie Python predefinite. È possibile installare librerie con ambito notebook dall'editor di notebook senza dover ricreare un cluster o ricollegare un notebook a un cluster. Le librerie con ambito notebook vengono create in un ambiente Python virtuale in modo da essere applicate esclusivamente alla sessione di notebook corrente. Ciò consente di isolare le dipendenze dei notebook. Per ulteriori informazioni, consulta [Utilizzo di librerie con ambito notebook](#) nella Guida alla gestione di Amazon EMR.
- EMRFS
 - È possibile abilitare una funzione di verifica ETag (Beta) impostando `fs.s3.consistent.metadata.etag.verification.enabled` su `true`. Con questa caratteristica, EMRFS utilizza gli ETag Amazon S3 per verificare che gli oggetti letti siano la versione più recente disponibile. Questa caratteristica è utile per i casi d'uso read-after-update in cui i file in Amazon S3 vengono sovrascritti mantenendo lo stesso nome. Questa caratteristica di verifica ETag attualmente non funziona con S3 Select. Per ulteriori informazioni, consulta [Configurazione della visualizzazione coerente](#).
- Spark
 - Le seguenti ottimizzazioni sono ora abilitate per impostazione predefinita: eliminazione dinamica delle partizioni, DISTINCT prima di INTERSECT, miglioramenti nell'inferenza delle statistiche del piano SQL per JOIN seguita da query DISTINCT, appiattimento delle sottoquery scalari, riordino di join ottimizzato e join del filtro Bloom. Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).
 - Migliorata la generazione del codice della fase per Sort Merge Join.
 - Migliorato il frammento di query e il riutilizzo delle sottoquery.
 - Miglioramenti per assegnare preventivamente gli executor all'avvio di Spark.
 - I join del filtro Bloom non vengono più applicati quando il lato più piccolo del join include un suggerimento per la trasmissione.
- Tez

- Risolto un problema con Tez. Ora l'interfaccia utente di Tez funziona su un cluster Amazon EMR con più nodi primari.

Problemi noti

- Le funzionalità migliorate di generazione del codice dell'intera fase per Sort Merge Join possono aumentare la pressione della memoria quando abilitate. Questa ottimizzazione migliora le prestazioni, ma può comportare tentativi multipli o errori di processo se `spark.yarn.executor.memoryOverheadFactor` non è regolato per fornire una memoria sufficiente. Per disabilitare questa funzionalità, imposta `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` su `false` (falso).
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.4	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.11.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.10.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.12.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.3.0	Connettore di EMR S3Select
<code>emrfs</code>	2.35.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.8.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.

Componente	Versione	Descrizione
hadoop-https-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.10	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.10	Servizio per servire una o più regioni HBase.
hbase-client	1.4.10	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.10	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.10	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.5-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.5-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.5-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.5-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.5-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.5-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.9.6	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.2-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.220	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.220	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.3	Client a riga di comando Spark.
spark-history-server	2.4.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.13.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.26.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.25.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
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	-	-	-	-

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.25.0. Le modifiche sono relative alla versione 5.24.1

Data del rilascio iniziale: 17 luglio 2019

Ultimo aggiornamento: 30 ottobre 2019

Amazon EMR 5.25.0

Aggiornamenti

- 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

Nuove funzionalità di

- (30 ottobre 2019) A partire da Amazon EMR versione 5.25.0, è possibile connettersi all'interfaccia utente di Spark History Server dalla pagina Summary (Riepilogo) del cluster o dalla scheda Application history (Cronologia applicazioni) nella console. Anziché configurare un proxy Web tramite una connessione SSH, puoi accedere rapidamente all'interfaccia utente del server della cronologia Spark per visualizzare i parametri dell'applicazione e accedere ai file di log pertinenti per cluster attivi e terminati. Per ulteriori informazioni, consulta [Accesso fuori cluster alle interfacce utente dell'applicazione persistente](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Spark
 - Migliorate le prestazioni di alcuni join utilizzando i filtri Bloom per pre-filtrare gli input. L'ottimizzazione è disattivata per impostazione predefinita e può essere abilitata impostando il parametro di configurazione Spark `spark.sql.bloomFilterJoin.enabled` su `true`.
 - Migliorate le prestazioni del raggruppamento per colonne di tipo stringa.
 - Migliorata la memoria dell'executor Spark predefinita e la configurazione dei nodi principali dei tipi di istanza R4 per i cluster senza HBase installato.
 - Risolto un problema precedente correlato alla caratteristica di eliminazione delle partizioni dinamiche in cui la tabella eliminata deve trovarsi sul lato sinistro del join.
 - Migliorata l'ottimizzazione di DISTINCT prima di INTERSECT da applicare a casi aggiuntivi che coinvolgono alias.

- Migliore inferenza delle statistiche del piano SQL per JOIN seguita da query DISTINCT. Questo miglioramento è disattivato per impostazione predefinita e può essere abilitato impostando il parametro di configurazione Spark `spark.sql.statsImprovements.enabled` su `true`. Questa ottimizzazione è richiesta dalla funzionalità Distinct prima di Intersect e verrà abilitata automaticamente quando `spark.sql.optimizer.distinctBeforeIntersect.enabled` è impostato su `true`.
- Ordine di join ottimizzato in base alle dimensioni della tabella e ai filtri. Questa ottimizzazione è disattivata per impostazione predefinita e può essere abilitata impostando il parametro di configurazione Spark `spark.sql.optimizer.sizeBasedJoinReorder.enabled` su `true`.

Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).

- EMRFS

- L'impostazione EMRFS, `fs.s3.buckets.create.enabled`, è ora disabilitata per impostazione predefinita. A seguito di alcuni test, abbiamo scoperto che la disattivazione di questa impostazione migliora le prestazioni e impedisce la creazione involontaria di bucket S3. Se l'applicazione si basa su questa funzionalità, è possibile attivarla impostando la proprietà `fs.s3.buckets.create.enabled` su `true` nella classificazione di configurazione `emrfs-site`. Per informazioni, consulta [Specifica di una configurazione durante la creazione di un cluster](#).
- Miglioramenti della crittografia del disco locale e della crittografia S3 nelle configurazioni di sicurezza (5 agosto 2019)
 - Separate le impostazioni di crittografia Amazon S3 dalle impostazioni di crittografia del disco locale nella configurazione della protezione.
 - Aggiunta un'opzione per abilitare la crittografia EBS con la versione 5.24.0 e successive. Selezionando questa opzione, il volume del dispositivo di root viene crittografato oltre ai volumi di archiviazione. Le versioni precedenti richiedevano l'uso di un'AMI personalizzata per crittografare il volume del dispositivo di root.
 - Per ulteriori informazioni, consulta [Opzioni di crittografia](#) nella Guida alla gestione di Amazon EMR.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.4	SDK Spark Amazon SageMaker
emr-ddb	4.10.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.9.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.34.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.8.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.5-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.5-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.5-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.5-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.5-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.5-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.6	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.220	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.220	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.3	Client a riga di comando Spark.
spark-history-server	2.4.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.13.1	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.2	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.14	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.14	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.25.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
recordserver-env	Modifica dei valori nell'ambiente EMR RecordServer.
recordserver-conf	Modifica dei valori nel file server.properties di EMR RecordServer.
recordserver-log4j	Modifica dei valori nel file log4j.properties di EMR RecordServer.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.24.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0.5.0	0.5.0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.24.1. Le modifiche sono relative alla versione 5.24.0.

Data del rilascio iniziale: 26 giugno 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.1	SDK Spark Amazon SageMaker
emr-ddb	4.9.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.8.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.33.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.8.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.4-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.6	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.219	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.219	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.2	Client a riga di comando Spark.
spark-history-server	2.4.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.24.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.24.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
SDK AWS per 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

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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
Presto	0.219	0.215	0.215	0.215

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.24.0. Le modifiche sono relative alla versione 5.23.0.

Data del rilascio iniziale: 11 giugno 2019

Ultimo aggiornamento: 5 agosto 2019

Aggiornamenti

- 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
- Connettori e driver:

- DynamoDB Connector 4.9.0
- MariaDB Connector 2.4.1
- Driver JDBC Amazon Redshift 1.2.27.1051

Modifiche, miglioramenti e problemi risolti

- Spark
 - Aggiunta ottimizzazione per eliminare dinamicamente le partizioni. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.dynamicPartitionPruning.enabled` su `true`.
 - Migliorate le prestazioni delle query INTERSECT. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.optimizer.distinctBeforeIntersect.enabled` su `true`.
 - Aggiunta ottimizzazione per appiattare le sottoquery scalari con aggregati che utilizzano la stessa relazione. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` su `true`.
 - Migliorata la generazione del codice della fase.

Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).

- Miglioramenti della crittografia del disco locale e della crittografia S3 nelle configurazioni di sicurezza (5 agosto 2019)
 - Separate le impostazioni di crittografia Amazon S3 dalle impostazioni di crittografia del disco locale nella configurazione della protezione.
 - Aggiunta un'opzione per abilitare la crittografia EBS. Selezionando questa opzione, il volume del dispositivo di root viene crittografato oltre ai volumi di archiviazione. Le versioni precedenti richiedevano l'uso di un'AMI personalizzata per crittografare il volume del dispositivo di root.
 - Per ulteriori informazioni, consulta [Opzioni di crittografia](#) nella Guida alla gestione di Amazon EMR.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.1	SDK Spark Amazon SageMaker
emr-ddb	4.9.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.8.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.3.0	Connettore di EMR S3Select
emrfs	2.33.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.8.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-4	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.8.5-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.5-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.4-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.4.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.6	Server multi-utente per notebook Jupyter
livy-server	0.6.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.4.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.219	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.219	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.2	Client a riga di comando Spark.
spark-history-server	2.4.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.24.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.23.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
SDK AWS per 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

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.1	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
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

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Note di rilascio

Questo è un rilascio di patch. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Stato del supporto IMDS (Instance Metadata Service) V2: Amazon EMR 5.23.1, 5.27.1 e 5.32 o versioni successive utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Per altri rilasci EMR 5.x, la disattivazione di IMDSv1 causa un errore di avvio del cluster.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.8.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.7.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.2.0	Connettore di EMR S3Select
<code>emrfs</code>	2.32.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.7.1	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-3	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httpfs-server	2.8.5-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.5-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.4-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark

Componente	Versione	Descrizione
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.23.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.23.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
SDK AWS per 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

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
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

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.23.0. Le modifiche sono relative alla versione 5.22.0.

Data del rilascio iniziale: 1° aprile 2019

Ultimo aggiornamento: 30 aprile 2019

Aggiornamenti

- AWS SDK for Java 1.11.519

Nuove funzionalità di

- (30 aprile 2019) Con Amazon EMR rilascio 5.23.0 e successivi, è possibile avviare un cluster con tre nodi primari per supportare applicazioni a elevata disponibilità come YARN Resource Manager, HDFS NameNode, Spark, Hive e Ganglia. Con questa caratteristica, il nodo primario non rappresenta più un potenziale singolo punto di errore. Se uno dei nodi primari ha esito negativo, Amazon EMR esegue automaticamente il failover in un nodo primario in standby e sostituisce il nodo primario guasto con uno nuovo con le medesime operazioni di configurazione e di bootstrap. Per ulteriori informazioni, consulta la sezione [Plan and Configure Primary Nodes](#) (Pianificazione e configurazione dei nodi primari).

Problemi noti

- Interfaccia utente Tez (corretta in Amazon EMR rilascio 5.26.0)

L'interfaccia utente di Tez non funziona su un cluster EMR con più nodi primari.

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica

che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.1	SDK Spark Amazon SageMaker
emr-ddb	4.8.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.7.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.32.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.7.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-hdfs-journalnode	2.8.5-amzn-3	Servizio HDFS per gestire il giornale di registrazione del file system Hadoop su cluster HA.
hadoop-httfs-server	2.8.5-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.8.5-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.3.4-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.

Componente	Versione	Descrizione
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.23.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.22.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
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
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

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.22.0. Le modifiche sono relative alla versione 5.21.0.

Important

A partire da Amazon EMR rilascio 5.22.0, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste inviate ad Amazon S3. I rilasci precedenti di Amazon EMR impiegano AWS Signature Version 2 in alcuni casi, a meno che le note di rilascio indichino l'utilizzo esclusivo di Signature Version 4. Per ulteriori informazioni, consulta [Autenticazione delle richieste \(AWS Signature Version 4\)](#) e [Autenticazione delle richieste \(AWS Signature Version 2\)](#) nella Guida per gli sviluppatori di Amazon Simple Storage.

Data del rilascio iniziale: 20 marzo 2019

Aggiornamenti

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Connettori e driver:
 - DynamoDB Connector 4.8.0

- MariaDB Connector 2.2.6
- Driver JDBC Amazon Redshift 1.2.20.1043

Nuove funzionalità di

- Modificata la configurazione EBS predefinita per i tipi di istanza EC2 con archiviazione solo EBS. Quando crei un cluster utilizzando Amazon EMR rilascio 5.22.0 e successivi, la quantità di archiviazione EBS predefinita aumenta in base alle dimensioni dell'istanza. Inoltre, l'archiviazione aumentata viene divisa su più volumi, offrendo migliori prestazioni IOPS. Se desideri utilizzare una configurazione dello storage delle istanze EBS diversa, puoi specificarla al momento della creazione di un cluster EMR o aggiungere nodi a un cluster esistente. Per ulteriori informazioni sulla quantità di archiviazione e sul numero di volumi allocati per impostazione predefinita per ciascun tipo di istanza, consulta [Archiviazione EBS di default per istanze](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Spark
 - Introdotta una nuova proprietà di configurazione per Spark su YARN, `spark.yarn.executor.memoryOverheadFactor`. Il valore di questa proprietà è un fattore di scala che imposta il valore del sovraccarico della memoria su una percentuale di memoria dell'executor, con un minimo di 384 MB. Se l'overhead della memoria è impostato esplicitamente utilizzando `spark.yarn.executor.memoryOverhead`, questa proprietà non ha effetto. Il valore predefinito è `0.1875`, che rappresenta il 18,75%. Questo valore predefinito per Amazon EMR lascia più spazio nei container YARN per l'overhead della memoria dell'executor rispetto al 10% predefinito impostato internamente da Spark. L'impostazione predefinita di Amazon EMR del 18,75% ha mostrato un minor numero di errori relativi alla memoria nei benchmark TPC-DS.
 - Backport di [SPARK-26316](#) per migliorare le prestazioni.
- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker

Componente	Versione	Descrizione
emr-ddb	4.8.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.6.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.31.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.7.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.8.5-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.8.5-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.8.5-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.9	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.9	Servizio per servire una o più regioni HBase.
hbase-client	1.4.9	Client di riga di comando HBase.
hbase-rest-server	1.4.9	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.9	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.4-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	5.1.0	Client a riga di comando Oozie.
oozie-server	5.1.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1,4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.1-HBase-1,4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.

Componente	Versione	Descrizione
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.22.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.

Classificazioni	Descrizione
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.

Classificazioni	Descrizione
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.

Classificazioni	Descrizione
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.

Classificazioni	Descrizione
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.21.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.30.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.7.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.5-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.8	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.8	Servizio per servire una o più regioni HBase.
hbase-client	1.4.8	Client di riga di comando HBase.
hbase-rest-server	1.4.8	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.8	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.4-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.21.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.

Classificazioni	Descrizione
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.21.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.6.2	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.214	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.21.1. Le modifiche sono relative alla versione 5.21.0.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.2.0	Connettore di EMR S3Select
<code>emrfs</code>	2.30.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.7.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.8.5-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.5-amzn-1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.8.5-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.8	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.8	Servizio per servire una o più regioni HBase.
hbase-client	1.4.8	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.8	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.8	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.21.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.

Classificazioni	Descrizione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.21.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.7.0	1.6.2	1.6.2	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.7
HCatalog	2.3.4	2.3.4	2.3.4	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.21.0. Le modifiche sono relative alla versione 5.20.0.

Data del rilascio iniziale: 18 febbraio 2019

Ultimo aggiornamento: 3 aprile 2019

Aggiornamenti

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Nuove funzionalità di

- (3 aprile 2019) Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console Amazon EMR, la AWS Command Line Interface (AWS CLI) o il kit SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Modifiche, miglioramenti e problemi risolti

- Zeppelin
 - Backport per [ZEPPELIN-3878](#).

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Tez
 - Questo problema è stato risolto in Amazon EMR 5.22.0.

Quando ci si connette all'interfaccia utente Tez all'indirizzo `http://MasterDNS:8080/tez-ui` tramite una connessione SSH al nodo primario del cluster, viene visualizzato l'errore "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled (Operazione adattatore fallita: server Timeline (ATS) fuori portata. O è inattivo o CORS non è abilitato)" o le attività mostrano una dicitura N/A inaspettata.

Ciò è dovuto al fatto che l'interfaccia utente Tez effettua richieste al server Timeline YARN utilizzando `localhost` anziché il nome host del nodo primario. Come soluzione alternativa, è disponibile uno script da eseguire come operazione o fase di bootstrap. Lo script aggiorna il nome host nel file `Tez configs .env`. Per ulteriori informazioni e per la posizione dello script, consulta [Istruzioni di bootstrap](#).

- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.11.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.30.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.7.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.5-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.8	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.8	Servizio per servire una o più regioni HBase.
hbase-client	1.4.8	Client di riga di comando HBase.
hbase-rest-server	1.4.8	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.8	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.4-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.215	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.215	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.21.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.

Classificazioni	Descrizione
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.20.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.6.2	1.6.2	1.6.1	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.7	1.4.7
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

⚠ Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.29.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-1	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.5-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.8	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.8	Servizio per servire una o più regioni HBase.
hbase-client	1.4.8	Client di riga di comando HBase.
hbase-rest-server	1.4.8	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.8	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.4-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.214	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.214	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.20.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop.
hadoop-ssl-client	Modifica la configurazione client ssl hadoop.
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.

Classificazioni	Descrizione
hbase-metrics	Modificare valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.

Classificazioni	Descrizione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
presto-connector-tpcds	Modifica i valori nel file <code>tpcds.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.

Classificazioni	Descrizione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.20.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.6.2	1.6.1	1.6.1	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.7	1.4.7	1.4.7
HCatalog	2.3.4	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.4
Hive	2.3.4	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
JupyterHub	0.9.4	0.9.4	0.9.4	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.0	1.3.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.212	0.212	0.210
Spark	2.4.0	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.11.0	1.11.0	1.9.0
Tez	0.9.1	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.12

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.20.0. Le modifiche sono relative alla versione 5.19.0.

Data del rilascio iniziale: 18 dicembre 2018

Ultimo aggiornamento: 22 gennaio 2019

Aggiornamenti

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

Nuove funzionalità di

- (22 gennaio 2019) Kerberos in Amazon EMR è stato migliorato per supportare l'autenticazione delle entità principali da un KDC esterno. In questo modo, la gestione dei principali viene centralizzata perché più cluster possono condividere un unico server KDC esterno. Inoltre, il server KDC esterno può avere un trust tra realm con un dominio Active Directory. In questo modo, tutti i cluster possono autenticare i principali da Active Directory. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione Kerberos](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- AMI predefinita di Amazon Linux per Amazon EMR
 - Il pacchetto Python3 è stato aggiornato da python 3.4 a 3.6.
- Il committer ottimizzato S3 EMRFS
 - Il committer EMRFS ottimizzato per S3 è ora abilitato per impostazione predefinita, il che migliora le prestazioni di scrittura. Per ulteriori informazioni, consulta [Utilizzare il committer ottimizzato S3 EMRFS](#).
- Hive
 - Backport per [HIVE-16686](#).
- Glue con Spark e Hive

- In EMR 5.20.0 o versioni successive, l'eliminazione delle partizioni parallele viene abilitata automaticamente per Spark e Hive quando AWS Glue Data Catalog viene utilizzato come metastore. Questa modifica riduce significativamente i tempi di pianificazione delle query eseguendo più richieste in parallelo per recuperare le partizioni. Il numero totale di segmenti che è possibile eseguire contemporaneamente è compreso tra 1 e 10. Il valore di default è 5, che è l'impostazione consigliata. Puoi modificare questa impostazione specificando la proprietà `aws.glue.partition.num.segments` nella classificazione di configurazione `hive-site`. Se si verifica una limitazione, è possibile disattivare la funzione modificando il valore su 1. Per ulteriori informazioni, consulta [Struttura del segmento AWS Glue](#).

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)
- Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Tez

- Questo problema è stato risolto in Amazon EMR 5.22.0.

Quando ci si connette all'interfaccia utente Tez all'indirizzo `http://MasterDNS:8080/tez-ui` tramite una connessione SSH al nodo primario del cluster, viene visualizzato l'errore "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled (Operazione adattatore fallita: server Timeline (ATS) fuori portata. O è inattivo o CORS non è abilitato)" o le attività mostrano una dicitura N/A inaspettata.

Ciò è dovuto al fatto che l'interfaccia utente Tez effettua richieste al server Timeline YARN utilizzando `localhost` anziché il nome host del nodo primario. Come soluzione alternativa, è disponibile uno script da eseguire come operazione o fase di bootstrap. Lo script aggiorna il nome host nel file `Tez configs .env`. Per ulteriori informazioni e per la posizione dello script, consulta [Istruzioni di bootstrap](#).

- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.2.0	Connettore di EMR S3Select
emrfs	2.29.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.5-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.8	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.8	Servizio per servire una o più regioni HBase.
hbase-client	1.4.8	Client di riga di comando HBase.
hbase-rest-server	1.4.8	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.8	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.4-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.4-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.4-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.4-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.4-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.4-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.3.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.1	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.214	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.214	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.4.0	Client a riga di comando Spark.
spark-history-server	2.4.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.4.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.4.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.12.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.9.1	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.20.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file <code>jupyterhub_config.py</code> di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file <code>config.json</code> di Sparkmagic.
livy-conf	Modifica i valori nel file <code>livy.conf</code> di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy <code>log4j.properties</code> .
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file <code>mapred-site.xml</code> dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file <code>oozie-log4j.properties</code> di Oozie.
oozie-site	Modifica i valori nel file <code>oozie-site.xml</code> di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.

Classificazioni	Descrizione
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.19.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.3.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

⚠ Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.2.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.1	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.1.0	Connettore di EMR S3Select
emrfs	2.28.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.5-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-0	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.5-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.7	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.7	Servizio per servire una o più regioni HBase.
hbase-client	1.4.7	Client di riga di comando HBase.
hbase-rest-server	1.4.7	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.7	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.3-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.212	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.212	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Client a riga di comando Spark.
spark-history-server	2.3.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.2	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.19.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.

Classificazioni	Descrizione
hbase-metrics	Modificare valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.

Classificazioni	Descrizione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-memory	Modifica i valori nel file memory.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.

Classificazioni	Descrizione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.19.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.6.1	1.6.0	1.6.0	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.19.0. Le modifiche sono relative alla versione 5.18.0.

Data del rilascio iniziale: 7 novembre 2018

Ultimo aggiornamento: 19 novembre 2018

Aggiornamenti

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- ZooKeeper 3.4.13
- AWS SDK for Java 1.11.433

Nuove funzionalità di

- (19 novembre 2018) EMR Notebooks è un ambiente gestito basato su Jupyter Notebook. Supporta i kernel Spark magic per PySpark, Spark SQL, Spark, R e Scala. EMR Notebooks può essere utilizzato con cluster creati utilizzando Amazon EMR rilascio 5.18.0 e successivi. Per ulteriori informazioni, consulta [Utilizzo di EMR Notebooks](#) nella Guida alla gestione di Amazon EMR.
- Il committer EMRFS ottimizzato per S3 è disponibile durante la scrittura di file Parquet tramite Spark ed EMRFS. Questo committer migliora le prestazioni di scrittura. Per ulteriori informazioni, consulta [Utilizzare il committer ottimizzato S3 EMRFS](#).

Modifiche, miglioramenti e problemi risolti

- YARN
 - Modificata la logica che limita il processo master dell'applicazione all'esecuzione sui nodi principali. Questa funzionalità ora utilizza la caratteristica e le proprietà delle etichette dei nodi YARN nelle classificazioni di configurazione `yarn-site` e `capacity-scheduler`. Per informazioni, consultare, <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- AMI predefinita di Amazon Linux per Amazon EMR
 - `ruby18`, `php56` e `gcc48` non vengono più installati per impostazione predefinita. Se lo si desidera, è possibile installarli tramite `yum`.

- Il Ruby Gem `aws-java-sdk` non è più installato per impostazione predefinita. Se lo si desidera, è possibile installarlo tramite `gem install aws-sdk`. Possono anche essere installati componenti specifici. Ad esempio, `gem install aws-sdk-s3`.

Problemi noti

- EMR Notebooks: in alcune circostanze, con più editor di notebook aperti, l'editor del notebook potrebbe non essere in grado di connettersi al cluster. In questo caso, cancella i cookie del browser e riapri gli editor dei notebook.
- Parametro `ContainerPending` e scalabilità automatica di CloudWatch: (corretto in 5.20.0) Amazon EMR può emettere un valore negativo per `ContainerPending`. Se `ContainerPending` viene utilizzato in una regola di scalabilità automatica, quest'ultima avrà un comportamento imprevisto. Non utilizzare `ContainerPending` con la scalabilità automatica.
- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.2.0	SDK Spark Amazon SageMaker
emr-ddb	4.7.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.5.1	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.1.0	Connettore di EMR S3Select
emrfs	2.28.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.8.5-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.5-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.5-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.5-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.5-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.5-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.5-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.5-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.5-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.7	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.7	Servizio per servire una o più regioni HBase.
hbase-client	1.4.7	Client di riga di comando HBase.
hbase-rest-server	1.4.7	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.7	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.3-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.9.4	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.3.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.

Componente	Versione	Descrizione
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.212	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.212	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing

Componente	Versione	Descrizione
spark-client	2.3.2	Client a riga di comando Spark.
spark-history-server	2.3.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.11.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.13	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.13	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.19.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-memory	Modifica i valori nel file <code>memory.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-tpcds	Modifica i valori nel file tpcds.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.18.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.6.0	1.6.0	1.5.2	1.5.2

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
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

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.1.3	SDK Spark Amazon SageMaker
emr-ddb	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.5.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.1.0	Connettore di EMR S3Select
emrfs	2.27.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.4-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.4-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.4-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.4-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.7	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.7	Servizio per servire una o più regioni HBase.
hbase-client	1.4.7	Client di riga di comando HBase.
hbase-rest-server	1.4.7	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.7	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.3-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.

Componente	Versione	Descrizione
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.210	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.210	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing

Componente	Versione	Descrizione
spark-client	2.3.2	Client a riga di comando Spark.
spark-history-server	2.3.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.9.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.18.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.18.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.206	0.206	0.206
Spark	2.3.2	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.18.0. Le modifiche sono relative alla versione 5.17.0.

Data del rilascio iniziale: 24 ottobre 2018

Aggiornamenti

- Flink 1.6.0
- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Nuove funzionalità di

- A partire da Amazon EMR rilascio 5.18.0, è possibile utilizzare i repository degli artefatti di Amazon EMR per creare il codice di processo secondo le versioni esatte di librerie e dipendenze disponibili con rilasci Amazon EMR specifici. Per ulteriori informazioni, consulta [Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR](#).

Modifiche, miglioramenti e problemi risolti

- Hive
 - Aggiunto il supporto per S3 Select. Per ulteriori informazioni, consulta [Utilizzo di S3 Select con Hive per migliorare le prestazioni](#).
- Presto
 - Aggiunto il supporto per [S3 Select](#) Pushdown. Per ulteriori informazioni, consulta [Uso di pushdown S3 Select con Presto per migliorare le prestazioni](#).

- Spark
 - La configurazione predefinita log4j per Spark è stata modificata per eseguire i registri del container ogni ora per i processi di streaming Spark. Ciò consente di impedire l'eliminazione dei registri per i processi Spark Streaming con esecuzione prolungata.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.1.3	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emr-s3-select	1.1.0	Connettore di EMR S3Select
emrfs	2.27.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.6.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.4-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-1	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.4-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.4-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.7	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.7	Servizio per servire una o più regioni HBase.
hbase-client	1.4.7	Client di riga di comando HBase.
hbase-rest-server	1.4.7	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.7	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.3-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
nginx	1.12.1	nginx [motore x] è un server proxy inverso e HTTP
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.

Componente	Versione	Descrizione
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.210	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.210	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.2	Client a riga di comando Spark.
spark-history-server	2.3.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.2	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.9.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.8.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.18.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.

Classificazioni	Descrizione
hbase-metrics	Modificare valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.

Classificazioni	Descrizione
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.

Classificazioni	Descrizione
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.17.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.2	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.6	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.1.3	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.0.0	Connettore di EMR S3Select
<code>emrfs</code>	2.26.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.5.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.4-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.4-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.4-amzn-1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.8.4-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.6	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.6	Servizio per servire una o più regioni HBase.
hbase-client	1.4.6	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.6	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.6	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.206	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.206	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Client a riga di comando Spark.
spark-history-server	2.3.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.9.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.17.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.17.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
SDK AWS per 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

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.0	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
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

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.17.1. Le modifiche sono relative alla versione 5.17.0.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.1.3	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.5.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emr-s3-select</code>	1.0.0	Connettore di EMR S3Select
<code>emrfs</code>	2.26.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.5.2	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.8.4-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.4-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.4-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.8.4-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.6	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.6	Servizio per servire una o più regioni HBase.
hbase-client	1.4.6	Client di riga di comando HBase.
hbase-rest-server	1.4.6	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.6	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.3.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.206	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.206	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.

Componente	Versione	Descrizione
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Client a riga di comando Spark.
spark-history-server	2.3.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.9.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.17.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .

Classificazioni	Descrizione
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.17.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Hadoop	2.8.4	2.8.4	2.8.4	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.4.0
MXNet	1.2.0	1.2.0	1.2.0	1.1.0
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)	-	-	-	-

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.17.0. Le modifiche sono relative alla versione 5.16.0.

Data del rilascio iniziale: 30 agosto 2018

Aggiornamenti

- Flink 1.5.2
- HBase 1.4.6
- Presto 0.206

Nuove funzionalità di

- Aggiunto il supporto per Tensorflow. Per ulteriori informazioni, consulta [TensorFlow](#).

Modifiche, miglioramenti e problemi risolti

- JupyterHub
 - Aggiunto il supporto per la persistenza notebook in Amazon S3. Per ulteriori informazioni, consulta [Configurazione della persistenza per i notebook in Amazon S3](#).
- Spark
 - Aggiunto il supporto per [S3 Select](#). Per ulteriori informazioni, consulta [Usare S3 Select con Spark per migliorare le prestazioni delle query](#).
- Risolti i problemi con i parametri Cloudwatch e la funzione di scalabilità automatica in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0.

Problemi noti

- Quando si crea un cluster che utilizza Kerberos con Livy installato, Livy ha esito negativo e genera un errore a indicare che l'autenticazione semplice non è abilitata. Il riavvio del server Livy risolve il problema. Per ovviare al problema, è possibile aggiungere una fase durante la creazione del cluster che esegue `sudo restart livy-server` sul nodo primario.
- Se si utilizza un'AMI Amazon Linux personalizzata basata su un'AMI Amazon Linux con una data di creazione 2018-08-11, il server Oozie non si avvia. Se si utilizza Oozie, è possibile creare un'AMI personalizzata basata su un ID AMI Amazon Linux con una data di creazione diversa. Puoi utilizzare il comando AWS CLI seguente per restituire un elenco di ID immagine per tutte le AMI Amazon Linux HVM con una versione 2018.03, insieme alla data di rilascio, in modo da poter scegliere un'AMI Linux appropriata come base. Sostituisci `MyRegion` con l'identificatore della Regione, ad esempio `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
```

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.1.3	SDK Spark Amazon SageMaker

Componente	Versione	Descrizione
emr-ddb	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.5.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emr-s3-select	1.0.0	Connettore di EMR S3Select
emrfs	2.26.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.5.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.8.4-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.4-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.8.4-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.8.4-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.6	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.6	Servizio per servire una o più regioni HBase.
hbase-client	1.4.6	Client di riga di comando HBase.
hbase-rest-server	1.4.6	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.6	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.3-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.206	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.206	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Client a riga di comando Spark.
spark-history-server	2.3.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.3.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tensorflow	1.9.0	Libreria software open source TensorFlow per il calcolo numerico ad alte prestazioni.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.17.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file <code>jupyterhub_config.py</code> di JupyterHubs.
jupyter-s3-conf	Configura la persistenza di S3 del notebook Jupyter.
jupyter-sparkmagic-conf	Modifica i valori nel file <code>config.json</code> di Sparkmagic.
livy-conf	Modifica i valori nel file <code>livy.conf</code> di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy <code>log4j.properties</code> .
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file <code>mapred-site.xml</code> dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file <code>oozie-log4j.properties</code> di Oozie.
oozie-site	Modifica i valori nel file <code>oozie-site.xml</code> di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-password-authenticator	Modifica i valori nel file <code>password-authenticator.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.16.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.5.0	1.5.0	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.4
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	-	-	-	-

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
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
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

⚠ Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.1.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.25.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.5.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.4-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-0	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.4-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.4-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.4-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.4	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.4	Servizio per servire una o più regioni HBase.
hbase-client	1.4.4	Client di riga di comando HBase.
hbase-rest-server	1.4.4	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.4	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.203	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.203	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Client a riga di comando Spark.
spark-history-server	2.3.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.16.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.

Classificazioni	Descrizione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.

Classificazioni	Descrizione
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.16.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.5.0	1.4.2	1.4.2	1.4.2

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.4.0	0.4.0	0.4.0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0
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	-	-	-	-

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.16.0. Le modifiche sono relative alla versione 5.15.0.

Data del rilascio iniziale: 19 luglio 2018

Aggiornamenti

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0
- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA 9.2
- Driver JDBC Redshift 1.2.15.1025

Modifiche, miglioramenti e problemi risolti

- HBase
 - Backport per [HBASE-20723](#)

- Presto
 - Modifiche alla configurazione per supportare l'autenticazione LDAP. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione LDAP con Presto on Amazon EMR](#).
- Spark
 - Apache Spark versione 2.3.1, disponibile a partire da Amazon EMR rilascio 5.16.0, risolve [CVE-2018-8024](#) e [CVE-2018-1334](#). Si consiglia di eseguire la migrazione delle versioni precedenti di Spark a Spark 2.3.1 o versioni successive.

Problemi noti

- Questa versione non supporta i tipi di istanza c1.medium o m1.small. I cluster che utilizzano uno di questi tipi di istanza non vengono avviati. Come soluzione alternativa, specifica un tipo di istanza diverso o utilizza una versione di rilascio diversa.
- Quando si crea un cluster che utilizza Kerberos con Livy installato, Livy ha esito negativo e genera un errore a indicare che l'autenticazione semplice non è abilitata. Il riavvio del server Livy risolve il problema. Per ovviare al problema, è possibile aggiungere una fase durante la creazione del cluster che esegue `sudo restart livy-server` sul nodo primario.
- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.1.0	SDK Spark Amazon SageMaker
emr-ddb	4.6.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.25.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.5.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.8.4-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.4-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.4-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.4-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.8.4-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.4-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.4-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.4-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.4-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.4	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.4	Servizio per servire una o più regioni HBase.
hbase-client	1.4.4	Client di riga di comando HBase.
hbase-rest-server	1.4.4	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.4	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.3-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.5.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.2.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	9.2.88	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.14.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.203	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.203	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.1	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.3.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.16.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-password-authenticator	Modifica i valori nel file password-authenticator.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.

Classificazioni	Descrizione
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.15.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.24.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.4.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.3-amzn-1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.8.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.4	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.4	Servizio per servire una o più regioni HBase.
hbase-client	1.4.4	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.4	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.4	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.1.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.15.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.15.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Liviy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.15.0. Le modifiche sono relative alla versione 5.14.0.

Data del rilascio iniziale: 21 giugno 2018

Aggiornamenti

- Aggiornato a HBase 1.4.4
- Aggiornato a Hive 2.3.3
- Aggiornato a Hue 4.2.0
- Oozie aggiornato a 5.0.0
- Aggiornato a Zookeeper 3.4.12
- AWS SDK aggiornato alla versione 1.11.333

Modifiche, miglioramenti e problemi risolti

- Hive
 - Backport per [HIVE-18069](#).
- Hue
 - Hue aggiornato per autenticarsi correttamente con Livy quando Kerberos è abilitato. Livy è ora supportato quando si utilizza Kerberos con Amazon EMR.
- JupyterHub
 - JupyterHub aggiornato affinché Amazon EMR possa installare le librerie client LDAP per impostazione predefinita.
 - Corretto un errore nello script che genera certificati autofirmati.

Problemi noti

- Questa versione non supporta i tipi di istanza c1.medium o m1.small. I cluster che utilizzano uno di questi tipi di istanza non vengono avviati. Come soluzione alternativa, specifica un tipo di istanza diverso o utilizza una versione di rilascio diversa.
- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in

Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emrfs	2.24.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-https-server	2.8.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.4	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.4	Servizio per servire una o più regioni HBase.
hbase-client	1.4.4	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.4	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.4	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.3-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.3-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.3-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.3-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.3-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.3-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.2.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.1.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	5.0.0	Client a riga di comando Oozie.
oozie-server	5.0.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.12	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.12	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.15.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.14.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Liviy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.23.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.8.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.2	Servizio per servire una o più regioni HBase.
hbase-client	1.4.2	Client di riga di comando HBase.
hbase-rest-server	1.4.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.1.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.14.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
container-log4j	Modifica i valori nel file container-log4j.properties di Hadoop YARN.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session.properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.

Classificazioni	Descrizione
hbase-metrics	Modificare valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file <code>log4j2.properties</code> i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file <code>beeline-log4j2.properties</code> di Hive.
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.

Classificazioni	Descrizione
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.14.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
SDK AWS per Java	1.11.297	1.11.297	1.11.297	1.11.297

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.0.0	1.0.0
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

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.14.1. Le modifiche sono relative alla versione 5.14.0.

Data del rilascio iniziale: 17 ottobre 2018

Aggiornata l'AMI predefinita per Amazon EMR al fine di risolvere potenziali vulnerabilità di sicurezza.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.0.1	SDK Spark Amazon SageMaker
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.23.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.8.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.8.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.2	Servizio per servire una o più regioni HBase.
hbase-client	1.4.2	Client di riga di comando HBase.
hbase-rest-server	1.4.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.1.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.14.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>flink-log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>flink-log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>flink-log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server-ssl-hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client-ssl-hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-parquet-logging	Modifica i valori nel file <code>parquet-logging.properties</code> di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file <code>hive-exec-log4j2.properties</code> di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file <code>llap-daemon-log4j2.properties</code> di Hive.
hive-log4j2	Modifica i valori nel file <code>hive-log4j2.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive
hiveserver2-site	Modifica i valori nel file <code>hiveserver2-site.xml</code> di Hive Server2
hue-ini	Modifica i valori nel file <code>ini</code> di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file <code>httpfs-site.xml</code> di Hadoop.
hadoop-kms-acls	Modifica i valori nel file <code>kms-acls.xml</code> di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file <code>kms-log4j.properties</code> di Hadoop.
hadoop-kms-site	Modifica i valori nel file <code>kms-site.xml</code> di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file <code>jupyter_notebook_config.py</code> di Jupyter Notebook.

Classificazioni	Descrizione
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.

Classificazioni	Descrizione
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.14.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.188
Spark	2.3.0	2.3.0	2.3.0	2.2.1
Sqoop	1.4.7	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.14.0. Le modifiche sono relative alla versione 5.13.0.

Data del rilascio iniziale: 4 giugno 2018

Aggiornamenti

- Aggiornato ad Apache Flink 1.4.2
- Apache MXNet aggiornato alla versione 1.1.0
- Apache Sqoop aggiornato alla versione 1.4.7

Nuove funzionalità di

- Aggiunto il supporto JupyterHub. Per ulteriori informazioni, consulta [JupyterHub](#).

Modifiche, miglioramenti e problemi risolti

- EMRFS
 - La stringa userAgent nelle richieste ad Amazon S3 è stata aggiornata per contenere le informazioni sull'utente e sul gruppo dell'entità principale che richiama. Questa può essere utilizzata con i log AWS CloudTrail per un monitoraggio più completo delle richieste.
- HBase
 - Include [HBASE-20447](#), che risolve un problema che potrebbe causare errori di cache, in particolare in caso di Regioni divise.
- MXNet
 - Aggiunte librerie OpenCV.
- Spark
 - Quando Spark scrive i file Parquet in un percorso Amazon S3 utilizzando EMRFS, l'algoritmo FileOutputCommitter è stato aggiornato per utilizzare la versione 2 anziché la versione 1. In questo modo viene ridotta la quantità di ridenominazioni, con un conseguente miglioramento delle prestazioni delle applicazioni. Questa modifica non influisce su:
 - Applicazioni diverse da Spark.
 - Applicazioni che scrivono su altri file system, ad esempio HDFS (che utilizzano ancora la versione 1 di FileOutputCommitter).
 - Applicazioni che utilizzano altri formati di output, come testo o csv, che già utilizzano la scrittura diretta EMRFS.

Problemi noti

- JupyterHub
 - L'utilizzo delle classificazioni di configurazione per impostare JupyterHub e singoli notebook Jupyter quando si crea un cluster non è supportato. Modifica manualmente il file jupyterhub_config.py e i file jupyter_notebook_config.py per ogni utente. Per ulteriori informazioni, consulta [Configurazione di JupyterHub](#).

- JupyterHub non si avvia sui cluster all'interno di una sottorete privata, generando il messaggio `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'`. Questo problema è causato da un errore nello script che genera certificati autofirmati. Utilizza la seguente soluzione alternativa per generare certificati autofirmati. Tutti i comandi vengono eseguiti mentre si è connessi al nodo primario.

1. Copia lo script di generazione del certificato dal container al nodo primario:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Utilizza un editor di testo per intervenire sulla riga 23 modificando il nome di host pubblico nel nome di host locale come illustrato di seguito:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Esegui lo script per generare certificati autofirmati:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Sposta i file di certificato generati dallo script nella directory `/etc/jupyter/conf/`:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

È possibile `tail` il file `jupyter.log` per verificare che JupyterHub sia stato riavviato e stia restituendo un codice di risposta 200. Ad esempio:

```
tail -f /var/log/jupyter/jupyter.log
```

Questo dovrebbe restituire una risposta simile alla seguente:

```
# [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
```

- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.23.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.4.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.3-amzn-1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.8.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.2	Servizio per servire una o più regioni HBase.
hbase-client	1.4.2	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
jupyterhub	0.8.1	Server multi-utente per notebook Jupyter
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.1.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.7	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.

Componente	Versione	Descrizione
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.14.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
container-log4j	Modifica i valori nel file <code>container-log4j.properties</code> di Hadoop YARN.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
jupyter-notebook-conf	Modifica i valori nel file jupyter_notebook_config.py di Jupyter Notebook.
jupyter-hub-conf	Modifica i valori nel file jupyterhub_config.py di JupyterHubs.
jupyter-sparkmagic-conf	Modifica i valori nel file config.json di Sparkmagic.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.13.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
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
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.22.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.2	Servizio per servire una o più regioni HBase.
hbase-client	1.4.2	Client di riga di comando HBase.
hbase-rest-server	1.4.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.13.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione <code>flink-conf.yaml</code> .

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.13.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.13.0. Le modifiche sono relative alla versione 5.12.0.

Aggiornamenti

- Spark aggiornato a 2.3.0

- HBase aggiornato a 1.4.2
- Presto aggiornato a 0.194
- Aggiornato AWS SDK for Java a 1.11.297

Modifiche, miglioramenti e problemi risolti

- Hive
 - Backport per [HIVE-15436](#). API di Hive ottimizzate per restituire solo visualizzazioni.

Problemi noti

- MxNet non dispone attualmente di librerie OpenCV.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.10.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.22.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.4.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.2	Servizio per servire una o più regioni HBase.
hbase-client	1.4.2	Client di riga di comando HBase.
hbase-rest-server	1.4.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-2	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-2	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-2	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-2	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.2-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.194	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.194	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
r	3.4.1	The R Project for Statistical Computing
spark-client	2.3.0	Client a riga di comando Spark.
spark-history-server	2.3.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.3.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.3.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.13.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione <code>flink-conf.yaml</code> .

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.12.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
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	0.188	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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

⚠ Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.9.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.21.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.4.0	Servizio per servire una o più regioni HBase.
hbase-client	1.4.0	Client di riga di comando HBase.
hbase-rest-server	1.4.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.2-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.188	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.188	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.12.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.12.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
MXNet	1.0.0	1.0.0	1.0.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.2. Le modifiche sono relative alla versione 5.12.1.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0.1	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.9.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.21.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.0	Servizio per servire una o più regioni HBase.
hbase-client	1.4.0	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.4.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.188	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.188	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.

Componente	Versione	Descrizione
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.12.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop

Classificazioni	Descrizione
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.

Classificazioni	Descrizione
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.

Classificazioni	Descrizione
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.

Classificazioni	Descrizione
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-redshift	Modifica i valori nel file redshift.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.12.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
SDK AWS per Java	1.11.267	1.11.267	1.11.238	1.11.238

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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
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

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.1. Le modifiche sono relative alla versione 5.12.0.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato *myapp-component* con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come *2.2-amzn-2*.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.0.1	SDK Spark Amazon SageMaker
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.9.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.21.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.0	Servizio per servire una o più regioni HBase.
hbase-client	1.4.0	Client di riga di comando HBase.
hbase-rest-server	1.4.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.3.2-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.

Componente	Versione	Descrizione
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.188	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.188	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.12.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.

Classificazioni	Descrizione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.12.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.4.0	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
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
Trino (PrestoSQL)	-	-	-	-

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.0. Le modifiche sono relative alla versione 5.11.1.

Aggiornamenti

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. Per ulteriori informazioni, consulta il [Log delle modifiche AWS SDK per Java](#) su GitHub
- Hadoop 2.7.3 ⇒ 2.8.3. Per ulteriori informazioni, consulta [Rilasci di Apache Hadoop](#).
- Flink 1.3.2 ⇒ 1.4.0. Per ulteriori informazioni, consulta l'[Annuncio del rilascio di Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Per ulteriori informazioni, consulta l'[Annuncio del rilascio di HBase](#).
- Hue 4.0.1 ⇒ 4.1.0. Per ulteriori informazioni, consulta le [Note di rilascio](#).
- MxNet 0.12.0 ⇒ 1.0.0. Per ulteriori informazioni, consulta il [Log delle modifiche MXNet](#) su GitHub.
- Presto 0.187 ⇒ 0.188. Per ulteriori informazioni, consulta le [Note di rilascio](#).

Modifiche, miglioramenti e problemi risolti

- Hadoop
 - La proprietà `yarn.resourcemanager.decommissioning.timeout` è stata modificata in `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. È possibile utilizzare questa proprietà per personalizzare il dimensionamento verso il basso del cluster. Per maggiori informazioni, consulta [Dimensionamento verso il basso del cluster](#) nella Guida alla gestione di Amazon EMR.
 - L'interfaccia CLI di Hadoop ha aggiunto l'opzione `-d` per il comando `cp` (copia), che specifica la copia diretta. Questo comando può essere utilizzato per evitare di creare un file `.COPYING` intermedio, rendendo più veloce la copia dei dati in Amazon S3. Per ulteriori informazioni, consulta [HADOOP-12384](#).
- Pig

- È stato aggiunto `pig-env`, che semplifica la configurazione delle proprietà dell'ambiente Pig. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).
- Presto
 - Aggiunta la classificazione della configurazione `presto-connector-redshift`, che puoi utilizzare per configurare i valori nel file di configurazione `redshift.properties` di Presto. Per ulteriori informazioni, consulta [Connettore Redshift](#) nella documentazione di Presto e [Configurazione delle applicazioni](#).
 - Il supporto Presto per EMRFS è stato aggiunto ed è la configurazione predefinita. I rilasci precedenti di Amazon EMR utilizzavano `PrestoS3FileSystem`, che era l'unica opzione disponibile. Per ulteriori informazioni, consulta [Configurazione EMRFS e PrestoS3FileSystem](#).

Note

Se esegui una query sui dati sottostanti in Amazon S3 con Amazon EMR versione 5.12.0, possono verificarsi errori di Presto. Questo avviene perché Presto non riesce a raccogliere i valori di classificazione della configurazione da `emrfs-site.xml`. Come soluzione alternativa, crea una sottodirectory `emrfs` in `usr/lib/presto/plugin/hive-hadoop2/` e crea un collegamento simbolico in `usr/lib/presto/plugin/hive-hadoop2/emrfs` sul file `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` esistente. Quindi riavvia il processo `presto-server` (`sudo presto-server stop` seguito da `sudo presto-server start`).

- Spark
 - Backport di [SPARK-22036: talvolta la moltiplicazione BigDecimal restituisce null](#).

Problemi noti

- MXNet non include le librerie OpenCV.
- SparkR non è disponibile per i cluster creati utilizzando un'AMI personalizzata perché R non è installato per impostazione predefinita sui nodi del cluster.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I

pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.0.1	SDK Spark Amazon SageMaker
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.9.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.21.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.4.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecos

Componente	Versione	Descrizione
		istema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.8.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.8.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.8.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.8.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.8.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.8.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.8.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.8.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.4.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.4.0	Servizio per servire una o più regioni HBase.
hbase-client	1.4.0	Client di riga di comando HBase.
hbase-rest-server	1.4.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.4.0	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.2-amzn-1	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-1	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-1	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-1	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.1.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.

Componente	Versione	Descrizione
mxnet	1.0.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.1.85	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.13.0-HBase-1.4	Le librerie Phoenix per server e client
phoenix-query-server	4.13.0-HBase-1.4	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.188	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.188	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.12.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-env	Change values in the Pig environment.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.

Classificazioni	Descrizione
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-redshift	Modifica i valori nel file <code>redshift.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.11.4 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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	2.3.2	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	-	-	-	-

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.0	SDK Spark Amazon SageMaker
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.8.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.11.4

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.

Classificazioni	Descrizione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.11.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
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	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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.11.3. Le modifiche sono relative alla versione 5.11.2.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.8.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.11.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.11.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
SDK AWS per 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	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.11.2. Le modifiche sono relative alla versione 5.11.1.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.8.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-6	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.7.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.

Componente	Versione	Descrizione
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.11.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop

Classificazioni	Descrizione
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.

Classificazioni	Descrizione
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.

Classificazioni	Descrizione
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.11.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
SDK AWS per Java	1.11.238	1.11.238	1.11.221	1.11.221
Python	2.7, 3.4	2.7, 3.4	2.7, 3.4	2.7, 3.4
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
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	2.3.2	2.3.2	2.3.1	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.1	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
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	-	-	-	-

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
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	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.11.1. Le modifiche sono relative al rilascio di Amazon EMR 5.8.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
aws-sagemaker-spark-sdk	1.0	SDK Spark Amazon SageMaker
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.8.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-6	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.3.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda

Componente	Versione	Descrizione
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.11.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.

Classificazioni	Descrizione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.11.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
SDK AWS per 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	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
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	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.11.0. Le modifiche sono relative alla versione 5.10.0.

Aggiornamenti

- Hive 2.3.2
- Spark 2.2.1
- SDK per Java: 1.11.238

Nuove funzionalità di

- Spark
 - Aggiunta l'impostazione `spark.decommissioning.timeout.threshold` che migliora il comportamento di disattivazione di Spark quando si utilizzano istanze Spot. Per ulteriori informazioni, consulta [Configurazione del comportamento di disattivazione di nodi](#).
 - Aggiunto a Spark il componente `aws-sagemaker-spark-sdk`, che installa Amazon SageMaker Spark e le dipendenze associate per l'integrazione di Spark con [Amazon SageMaker](#). Puoi utilizzare Amazon SageMaker Spark per configurare le pipeline di machine learning (ML) di Spark utilizzando le fasi di Amazon SageMaker. Per ulteriori informazioni, consulta il [File readme di SageMaker Spark](#) su GitHub e [Utilizzare Apache Spark con Amazon SageMaker](#) nella Guida per lo sviluppatore di Amazon SageMaker.

Problemi noti

- MXNet non include le librerie OpenCV.
- Per impostazione predefinita, Hive 2.3.2 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>aws-sagemaker-spark-sdk</code>	1.0	SDK Spark Amazon SageMaker
<code>emr-ddb</code>	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.8.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-6	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-6	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-6	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-6	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-6	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.7.3-amzn-6	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-6	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-6	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.2-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.2-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.2-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.2-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.2-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.

Componente	Versione	Descrizione
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.1	Client a riga di comando Spark.
spark-history-server	2.2.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.

Componente	Versione	Descrizione
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.11.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop

Classificazioni	Descrizione
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.

Classificazioni	Descrizione
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.

Classificazioni	Descrizione
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.

Classificazioni	Descrizione
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.10.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
SDK AWS per Java	1.11.221	1.11.221	1.11.183	1.11.183
Python	2.7, 3.4	2.7, 3.4	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
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	2.3.1	2.3.1	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.1	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.7.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.3-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.3-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.10.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.

Classificazioni	Descrizione
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.

Classificazioni	Descrizione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.10.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
SDK AWS per Java	1.11.221	1.11.183	1.11.183	1.11.160
Python	2.7, 3.4	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	-
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.16.0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.10.0. Le modifiche sono relative al rilascio di Amazon EMR 5.9.0.

Aggiornamenti

- AWS SDK for Java 1.11.221

- Hive 2.3.1
- Presto 0.187

Nuove funzionalità di

- Aggiunto supporto per l'autenticazione Kerberos. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione Kerberos](#) nella Guida alla gestione di Amazon EMR
- Aggiunto supporto per i ruoli IAM per EMRFS. Per ulteriori informazioni, consulta [Configurazione dei ruoli IAM per le richieste EMRFS ad Amazon S3](#) nella Guida alla gestione di Amazon EMR
- Aggiunto supporto per i tipi di istanza P2 e P3 basati su GPU. Per ulteriori informazioni, consulta [Istanze P2 di Amazon EC2](#) e [Istanze P3 di Amazon EC2](#). NVIDIA driver 384.81 e CUDA driver 9.0.176 sono installati su questi tipi di istanza per impostazione predefinita.
- Aggiunta del supporto per [Apache MXNet](#).

Modifiche, miglioramenti e problemi risolti

- Presto
 - Aggiunto supporto per l'utilizzo di AWS Glue Data Catalog come metastore predefinito di Hive. Per ulteriori informazioni, consulta [Utilizzo di Presto con AWS Glue Data Catalog](#).
 - Aggiunto supporto per [funzioni geospaziali](#).
 - Aggiunto supporto [spill to disk](#) per join.
 - Aggiunto supporto per il [connettore Redshift](#).
- Spark
 - Backport per [SPARK-20640](#), che rende il timeout rcp e i tentativi per i valori di registrazione shuffle configurabili utilizzando le `spark.shuffle.registration.timeout` e `spark.shuffle.registration.maxAttempts` proprietà.
 - Backport per [SPARK-21549](#), che corregge un errore che avviene durante la scrittura di `OutputFormat` personalizzato su percorsi non-HDFS.
 - Backport per [Hadoop-13270](#)
- Le librerie Numpy, Scipy e Matplotlib sono state rimosse dall'AMI di base di Amazon EMR. Se hai bisogno di queste librerie per tua applicazione, sono disponibili nella repository dell'applicazione, così potrai usare un'operazione di bootstrap per installarle su tutti i nodi usando `yum install`.

- L'AMI di base Amazon EMR non dispone più dei pacchetti RPM di applicazione inclusi, per cui i pacchetti RPM non sono più presenti sui nodi cluster. Adesso, le AMI personalizzate e l'AMI di base Amazon EMR fanno riferimento al repository dei pacchetti RPM in Amazon S3.
- A causa dell'introduzione della fatturazione al secondo in Amazon EC2, il valore predefinito Scale down behavior (Comportamento di dimensionamento verso il basso) è diventato Terminate at task completion (Termina al completamento dell'attività) anziché Terminate at instance hour (Termina all'ora dell'istanza). Per maggiori informazioni, vedi [Configura il ridimensionamento del cluster](#).

Problemi noti

- MXNet non include le librerie OpenCV.
- Per impostazione predefinita, Hive 2.3.1 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.5.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.4.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.7.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.20.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.3-amzn-5	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-5	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-5	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-5	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-5	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-5	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-5	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-5	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.3-amzn-5	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.3.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mxnet	0.12.0	Una libreria flessibile, scalabile ed efficiente per il deep learning.
mysql-server	5.5.54+	Server di database MySQL.
nvidia-cuda	9.0.176	Driver Nvidia e kit di strumenti Cuda
oozie-client	4.3.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.187	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.187	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.3	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.10.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.

Classificazioni	Descrizione
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.

Classificazioni	Descrizione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.9.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questa versione sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
SDK AWS per Java	1.11.183	1.11.183	1.11.160	1.11.160
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.16.0	0.16.0
Presto	0.184	0.184	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

⚠ Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.7.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emrfs	2.19.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.2	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-https-server	2.7.3-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop

Componente	Versione	Descrizione
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.184	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.184	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.0	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.9.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.

Classificazioni	Descrizione
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.9.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questa versione sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
SDK AWS per Java	1.11.183	1.11.160	1.11.160	1.11.160
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
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
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)	-	-	-	-

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.9.0. Le modifiche sono relative al rilascio di Amazon EMR 5.8.0.

Data del rilascio: 5 ottobre 2017

Ultimo aggiornamento della caratteristica: 12 ottobre 2017

Aggiornamenti

- AWS SDK for Java versione 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Nuove funzionalità di

- Aggiunto il supporto per Livy (versione 0.4.0-incubazione). Per ulteriori informazioni, consulta [Apache Livy](#).
- Aggiunto il supporto per Hue Notebook per Spark.
- Aggiunto supporto per le istanze Amazon EC2 serie i3 (12 ottobre 2017).

Modifiche, miglioramenti e problemi risolti

- Spark
 - Aggiunto un nuovo set di caratteristiche per garantire che Spark gestisca più elegantemente la terminazione dei nodi a causa di un ridimensionamento manuale o di una richiesta di policy di scalabilità automatica. Per ulteriori informazioni, consulta [Configurazione del comportamento di disattivazione di nodi](#).

- Viene utilizzato SSL invece di 3DES per la crittografia in transito del servizio di trasferimento dei blocchi, migliorando così le prestazioni dei tipi di istanza Amazon EC2 con AES-NI.
- Backport per [SPARK-21494](#).
- Zeppelin
 - Backport per [ZEPPELIN-2377](#).
- HBase
 - Aggiunta patch [HBASE-18533](#) che permette valori aggiuntivi per la configurazione HBase BucketCache utilizzando la classificazione della configurazione `hbase-site`.
- Hue
 - Aggiunto il supporto AWS Glue Data Catalog per l'editor delle query Hive su Hue.
 - Per impostazione predefinita, i superuser su Hue possono accedere a tutti i file a cui hanno accesso i ruoli IAM di Amazon EMR. Gli utenti appena creati non dispongono automaticamente delle autorizzazioni per accedere al filebrowser Amazon S3 e devono disporre delle autorizzazioni `filebrowser.s3_access` attivate per il proprio gruppo.
- Risolto un problema che causava la non accessibilità ai dati JSON sottostanti creati utilizzando AWS Glue Data Catalog.

Problemi noti

- L'avvio del cluster non va a buon fine quando tutte le applicazioni sono installate e la dimensione del volume di root Amazon EBS di default non viene modificata. Per risolvere il problema, utilizza il comando `aws emr create-cluster` da AWS CLI e specifica un parametro `--ebs-root-volume-size` più grande.
- Per impostazione predefinita, Hive 2.3.0 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.7.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.19.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.3.2	Applicazioni e script client a riga di comando Apache Flink.

Componente	Versione	Descrizione
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.7.3-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	4.0.1	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
livy-server	0.4.0-incubating	Interfaccia REST per l'interazione con Apache Spark
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.

Componente	Versione	Descrizione
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.184	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.184	Servizio per l'esecuzione di parti di una query.
pig-client	0.17.0	Client a riga di comando Pig.
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.9.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.

Classificazioni	Descrizione
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
livy-conf	Modifica i valori nel file livy.conf di Livy.
livy-env	Modifica i valori nell'ambiente Livy.
livy-log4j	Modifica le impostazioni Livy log4j.properties.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.8.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
SDK AWS per Java	1.11.160	1.11.160	1.11.160	1.11.160
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.6.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-3	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.7.3-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-1	Client a riga di comando Pig.

Componente	Versione	Descrizione
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.8.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.8.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
SDK AWS per Java	1.11.160	1.11.160	1.11.160	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1
Hudi	-	-	-	-

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.8.2. Le modifiche sono relative alla versione 5.8.1.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.6.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-3	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.7.3-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-1	Client a riga di comando Pig.

Componente	Versione	Descrizione
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.8.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.8.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
SDK AWS per Java	1.11.160	1.11.160	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.8.1. Le modifiche sono relative al rilascio di Amazon EMR 5.8.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.6.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.3-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-1	Client a riga di comando Pig.
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.8.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.8.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
SDK AWS per Java	1.11.160	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Delta	-	-	-	-
Flink	1.3.1	1.3.0	1.3.0	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.0
HCatalog	2.3.0	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
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

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.8.0. Le modifiche sono relative al rilascio di Amazon EMR 5.7.0.

Data del rilascio iniziale: 10 agosto 2017

Ultimo aggiornamento della caratteristica: 25 settembre 2017

Aggiornamenti

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Per ulteriori informazioni, consulta [Note di rilascio](#) sul sito di Apache Hive.
- Spark 2.2.0. Per ulteriori informazioni, consulta [Note di rilascio](#) sul sito di Apache Spark.

Nuove funzionalità di

- Aggiunto supporto per la visualizzazione della cronologia dell'applicazione (25 settembre 2017). Per ulteriori informazioni, consulta [Visualizzazione della cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Integrazione con Glue Data Catalog di AWS
 - Aggiunta funzionalità che permette a Hive e Spark SQL di utilizzare AWS Glue Data Catalog come archivio dei metadati di Hive. Per ulteriori informazioni, consulta [Utilizzo di AWS Glue Data Catalog come metastore Hive](#) e [Utilizzare AWS Glue Data Catalog come metastore Spark SQL](#).
- Aggiunto Application history (Cronologia dell'applicazione) ai dettagli del cluster, il che ti permette di visualizzare i dati cronologici delle applicazioni YARN e i dettagli aggiuntivi per le applicazioni Spark. Per ulteriori informazioni, consulta [Visualizzazione della cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.
- Oozie
 - Backport per [OOZIE-2748](#).
- Hue
 - Backport per [HUE-5859](#)
- HBase
 - Aggiunta patch per esporre il tempo di avvio del server master HBase tramite le Java Management Extensions (JMX) utilizzando `getMasterInitializedTime`.
 - Aggiunta patch che migliora il tempo di avvio del cluster.

Problemi noti

- L'avvio del cluster non va a buon fine quando tutte le applicazioni sono installate e la dimensione del volume di root Amazon EBS di default non viene modificata. Per risolvere il problema, utilizza il comando `aws emr create-cluster` da AWS CLI e specifica un parametro `--ebs-root-volume-size` più grande.
- Per impostazione predefinita, Hive 2.3.0 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

- Spark: quando si utilizza Spark, c'è un problema di perdita del gestore di file con il daemon appusher, che può apparire per un processo Spark a esecuzione prolungata dopo diverse ore o giorni. Per risolvere il problema, connettiti al nodo primario e digita `sudo /etc/init.d/appusher stop`. Questa azione arresta l'appusher daemon, che Amazon EMR riavvierà automaticamente.
- Cronologia dell'applicazione
 - I dati cronologici per gli executor Spark terminati non sono disponibili.
 - La cronologia dell'applicazione non è disponibile per i cluster che utilizzano una configurazione di sicurezza per attivare la crittografia in-flight.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.4.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.4.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.4.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.6.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-3	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.7.3-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.3.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.3.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.3.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.3.0-amzn-0	Client a riga di comando Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.3.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-1	Client a riga di comando Pig.

Componente	Versione	Descrizione
spark-client	2.2.0	Client a riga di comando Spark.
spark-history-server	2.2.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.2.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.2.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.8.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.7.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.

Componente	Versione	Descrizione
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.

Componente	Versione	Descrizione
spark-client	2.1.1	Client a riga di comando Spark.
spark-history-server	2.1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.7.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione <code>server ssl hadoop</code> .
hadoop-ssl-client	Modifica la configurazione <code>client ssl hadoop</code> .
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.

Classificazioni	Descrizione
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.7.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.7.0. Le modifiche sono relative al rilascio di Amazon EMR 5.6.0.

Data del rilascio: 13 luglio 2017

Aggiornamenti

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Nuove funzionalità di

- Aggiunta la capacità di specificare un'AMI Amazon Linux quando si crea un cluster. Per ulteriori informazioni, consulta [Utilizzo di un AMI personalizzato](#).

Modifiche, miglioramenti e problemi risolti

- HBase
 - Aggiunta la capacità di configurare la replica di lettura ai cluster di HBase. Consulta [Utilizzare la replica di lettura al cluster](#).
 - Numerosi miglioramenti e correzioni dei bug.
- Presto: aggiunta la possibilità di configurare `node.properties`.
- YARN: aggiunta la possibilità di configurare `container-log4j.properties`.
- Sqoop: backport per [SQOOP-2880](#), che introduce un argomento che ti permette di impostare la directory temporanea di Sqoop.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.18.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.3.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.11.0-HBase-1.3	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.11.0-HBase-1.3	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.1	Client a riga di comando Spark.
spark-history-server	2.1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.7.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-env	Modifica i valori nel file <code>presto-env.sh</code> di Presto.
presto-node	Modifica i valori nel file <code>node.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.
sqoop-site	Modifica i valori nel file <code>sqoop-site.xml</code> di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.6.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.1	1.2.1	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.12.2	0.12.2

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.2.1	Applicazioni e script client a riga di comando Apache Flink.

Componente	Versione	Descrizione
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.1	Client a riga di comando Spark.
spark-history-server	2.1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.1	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.6.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive

Classificazioni	Descrizione
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.

Classificazioni	Descrizione
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Spark
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file <code>oraoop-site.xml</code> di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.6.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.6.0. Le modifiche sono relative al rilascio di Amazon EMR 5.5.0.

Data del rilascio: 5 giugno 2017

Aggiornamenti

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. Questa è la prima versione di Mahout che supporta Spark 2.x in Amazon EMR versione 5.0 e successive.

- Spark 2.1.1

Modifiche, miglioramenti e problemi risolti

- Presto
 - Aggiunta la capacità di attivare la comunicazione sicura SSL/TLS tra i nodi Presto attivando la cifratura in transito tramite la configurazione di sicurezza. Per maggiori informazioni, consulta [Cifratura dei dati in transito](#).
 - Backport per [Presto 7661](#) che aggiunge l'opzione VERBOSE all'istruzione EXPLAIN ANALYZE per comunicare statistiche più dettagliate di basso livello relative a un piano di query.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.1	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.3.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.1	Servizio per servire una o più regioni HBase.
hbase-client	1.3.1	Client di riga di comando HBase.
hbase-rest-server	1.3.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.13.0	Librerie per Machine Learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.1	Client a riga di comando Spark.
spark-history-server	2.1.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.6.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modificare valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-env	Modifica i valori nel file presto-env.sh di Presto.
presto-node	Modifica i valori nel file node.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.5.4 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.16.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.5.4

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.5.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.5.3. Le modifiche sono relative alla versione 5.5.2.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.16.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccogliitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
<code>ganglia-web</code>	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.5.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.5.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.166

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
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	0.7.1	0.7.1	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.9

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.5.2. Le modifiche sono relative alla versione 5.5.1.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.3.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.16.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
<code>ganglia-web</code>	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.5.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.5.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.166	0.166

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.5.1. Le modifiche sono relative al rilascio di Amazon EMR 5.5.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.16.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.5.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.5.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.166	0.166	0.157.1

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.5.0. Le modifiche sono relative al rilascio di Amazon EMR 5.4.0.

Data del rilascio: 26 aprile 2017

Aggiornamenti

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Modifiche, miglioramenti e problemi risolti

- Spark
 - Backport per Spark Patch [\(SPARK-20115\) Fix DAGScheduler per ricalcolare tutti i blocchi shuffle perduti durante la non disponibilità del servizio shuffle esterno](#) alla versione 2.1.0 di Spark, che è inclusa in questo rilascio.
- Flink

- Flink ora è integrato con Scala 2.11. Se utilizzi le API e le librerie Scala, ti consigliamo di utilizzare Scala 2.11 nei tuoi progetti.
- Gestito un problema in cui i valori predefiniti di `HADOOP_CONF_DIR` e `YARN_CONF_DIR` non erano impostati correttamente, per cui `start-scala-shell.sh` non funzionava. Aggiunta inoltre la capacità di impostare tali valori usando `env.hadoop.conf.dir` e `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` o la classificazione della configurazione `flink-conf`.
- Introdotta un nuovo comando specifico per EMR, `flink-scala-shell` come wrapper per `start-scala-shell.sh`. Ti consigliamo di usare questo comando anziché `start-scala-shell`. Il nuovo comando semplifica l'esecuzione. Ad esempio, `flink-scala-shell -n 2` avvia uno shell Flink Scala con un parallelismo dell'attività di 2.
- Introdotta un nuovo comando specifico per EMR, `flink-yarn-session` come wrapper per `yarn-session.sh`. Ti consigliamo di usare questo comando anziché `yarn-session`. Il nuovo comando semplifica l'esecuzione. Ad esempio, `flink-yarn-session -d -n 2` avvia una sessione Flink di lunga durata in uno stato scollegato con due task manager.
- Risolto [\(FLINK-6125\) Commons httpclient non più schermato in Flink 1.2](#)
- Presto
 - Aggiunto supporto per l'autenticazione LDAP. L'utilizzo di LDAP con Presto on Amazon EMR richiede di abilitare l'accesso HTTPS per il coordinatore di Presto (`http-server.https.enabled=true` in `config.properties`). Per dettagli sulla configurazione, consulta [Autenticazione LDAP](#) nella documentazione di Presto.
 - Aggiunta del supporto per `SHOW GRANTS`.
- AMI Linux di base di Amazon EMR
 - I rilasci di Amazon EMR sono ora basati su Amazon Linux 2017.03. Per ulteriori informazioni, consulta la sezione relativa alle [note di rilascio dell'AMI Amazon Linux 2017.03](#).
 - È stato rimosso Python 2.6 dall'immagine Linux di base Amazon EMR. Python 2.7 e 3.4 vengono installati per impostazione predefinita. Puoi installare Python 2.6 manualmente se necessario.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni

più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.5.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.16.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.12.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.170	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.170	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.10	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.10	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.5.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione <code>flink-conf.yaml</code> .

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.4.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.166	0.157.1	0.157.1

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.3.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.15.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.166	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.166	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.0	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.4.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.4.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.2.0	1.1.4	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.157.1	0.157.1	0.157.1

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.4.0. Le modifiche sono relative al rilascio di Amazon EMR 5.3.0.

Data di rilascio: 08 marzo 2017

Aggiornamenti

- Aggiornato a Flink 1.2.0
- Aggiornato a HBase 1.3.0
- Aggiornato a Phoenix 4.9.0

Note

Se effettui l'aggiornamento da una versione precedente di Amazon EMR ad Amazon EMR versione 5.4.0 o successive e utilizzi un'indicizzazione secondaria, effettua l'aggiornamento degli indici locali come descritto nella documentazione su [Apache Phoenix](#). Amazon EMR rimuove le configurazioni richieste dalla classificazione `hbase-site`, ma gli indici devono essere ripopolati. Gli aggiornamenti online e offline degli indici sono supportati. Gli aggiornamenti online sono quelli predefiniti, il che significa che gli indici vengono ripopolati durante la fase di inizializzazione dai client Phoenix della versione

4.8.0 o più grande. Per specificare gli aggiornamenti offline, imposta la configurazione `phoenix.client.localIndexUpgrade` su `false` nella classificazione `phoenix-site` e poi SSH sul nodo primario per eseguire `psql [zookeeper] -1`.

- Aggiornato a Presto 0.166
- Aggiornato a Zeppelin 0.7.0

Modifiche e miglioramenti

- Aggiunto supporto per le istanze r4. Consulta [Tipi di istanza Amazon EC2](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.3.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.15.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.2.0	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.3.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.3.0	Servizio per servire una o più regioni HBase.
hbase-client	1.3.0	Client di riga di comando HBase.
hbase-rest-server	1.3.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.3.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.

Componente	Versione	Descrizione
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server2	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.9.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.9.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.166	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.166	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.7.0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.4.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.3.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.14.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.4	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.3.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.3.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
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

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.3.1. Le modifiche sono relative al rilascio di Amazon EMR 5.3.0.

Data di rilascio: 7 febbraio 2017

Modifiche minori alle patch Zeppelin di backport e aggiornamento dell'AMI predefinita per Amazon EMR.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.14.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.4	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.3.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.3.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.1.4	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
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	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.3.0. Le modifiche sono relative al rilascio di Amazon EMR 5.2.1.

Data del rilascio: 26 gennaio 2017

Aggiornamenti

- Aggiornato a Hive 2.1.1
- Aggiornato a Hue 3.11.0
- Aggiornato a Spark 2.1.0
- Aggiornato a Oozie a 4.3.0
- Aggiornato a Flink 1.1.4

Modifiche e miglioramenti

- Aggiunta una patch a Hue che ti permette di utilizzare l'impostazione `interpreters_shown_on_wheel` per configurare che interpreti mostrare per primi sulla ruota di selezione Notebook, indipendentemente dall'ordine nel file `hue.ini`.
- Aggiunta la classificazione della configurazione `hive-parquet-logging`, che puoi utilizzare per configurare i valori nel file `parquet-logging.properties` di Hive.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR

e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.14.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.1.4	Applicazioni e script client a riga di comando Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.1-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.1.1-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.1-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.1-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.1-amzn-0	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.1-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.11.0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.3.0	Client a riga di comando Oozie.
oozie-server	4.3.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.1.0	Client a riga di comando Spark.
spark-history-server	2.1.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.1.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.1.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.3.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione <code>flink-conf.yaml</code> .

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-parquet-logging	Modifica i valori nel file parquet-logging.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.2.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.152.3

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Spark	2.0.2	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.13.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.3	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.2	Client a riga di comando Spark.
spark-history-server	2.0.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.2.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.2.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
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	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.2. Le modifiche sono relative al rilascio di Amazon EMR 5.2.1.

Data di rilascio: 2 maggio 2017

Problemi noti risolti dai rilasci precedenti

- Backport per [SPARK-194459](#), che risolve un problema in cui la lettura da una tabella ORC con colonne char/varchar restituisce un errore.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.13.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.3	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.2	Client a riga di comando Spark.
spark-history-server	2.0.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.2.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.2.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.1	2.0.1

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
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	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.8	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.1. Le modifiche sono relative al rilascio di Amazon EMR 5.2.0.

Data del rilascio: 29 dicembre 2016

Aggiornamenti

- Aggiornato a Presto 0.157.1. Per ulteriori informazioni, consulta la sezione delle [note di rilascio di Presto](#) nella documentazione di Presto.
- Aggiornato a Zookeeper 3.4.9. Per ulteriori informazioni, consulta [Note di rilascio di ZooKeeper](#) nella documentazione di Apache ZooKeeper.

Modifiche e miglioramenti

- Aggiunto supporto per il tipo di istanza Amazon EC2 m4.16xlarge in Amazon EMR versione 4.8.3 e successive, escluse 5.0.0, 5.0.3 e 5.2.0.
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- La posizione dei percorsi di configurazione Flink e YARN è ora impostata per default in `/etc/default/flink` e non occorre impostare le variabili di ambiente `FLINK_CONF_DIR` e

HADOOP_CONF_DIR quando si eseguono gli script dei driver `flink` o `yarn-session.sh` per avviare processi Flink.

- Aggiunto supporto per la classe `FlinkKinesisConsumer`.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in Hadoop in cui il thread `ReplicationMonitor` poteva bloccarsi per molto tempo a causa di un conflitto tra la replica e la cancellazione dello stesso file in un cluster di grandi dimensioni.
- Risolto un problema in cui `ControlledJob#toString` restituiva un errore con una `null pointer exception` (NPE) quando lo stato del processo non era aggiornato correttamente.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.13.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.3	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.2	Client a riga di comando Spark.
spark-history-server	2.0.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.2.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .
flink-log4j-yarn-session	Modifica le impostazioni <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Modifica le impostazioni <code>Flink log4j-cli.properties</code> .

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.

Classificazioni	Descrizione
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.

Classificazioni	Descrizione
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.2.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.1	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.0. Le modifiche sono relative al rilascio di Amazon EMR 5.1.0.

Data del rilascio: 21 novembre 2016

Modifiche e miglioramenti

- Aggiunta la modalità di archiviazione Amazon S3 per HBase.
- Ti permette di specificare un percorso Amazon S3 per la rootdir HBase. Per ulteriori informazioni, consulta [HBase su Amazon S3](#).

Aggiornamenti

- Aggiornato a Spark 2.0.2

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con /mnt vincolato a 2 TB sui tipi di istanza solo EBS.
- Risolto un problema con il controller dell'istanza e i log logpusher che venivano indirizzati ai file .out corrispondenti anziché ai normali file .log log4j-configured, che ruotano ogni ora. I file .out non ruotano e questo determinerà il riempimento della partizione /emr. Questo problema interessa soltanto i tipi di istanza hardware virtual machine (HVM).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.1.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.12.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>flink-client</code>	1.1.3	Applicazioni e script client a riga di comando Apache Flink.

Componente	Versione	Descrizione
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.

Componente	Versione	Descrizione
hadoop-mapred	2.7.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.

Componente	Versione	Descrizione
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.

Componente	Versione	Descrizione
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.152.3	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.152.3	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.2	Client a riga di comando Spark.
spark-history-server	2.0.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.2	Librerie Apache Spark necessarie per gli slave YARN.

Componente	Versione	Descrizione
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.2.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione flink-conf.yaml.
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase	Impostazioni Amazon EMR per Apache HBase.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.

Classificazioni	Descrizione
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2

Classificazioni	Descrizione
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.

Classificazioni	Descrizione
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-blackhole	Modifica i valori nel file <code>blackhole.properties</code> di Presto.
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.

Classificazioni	Descrizione
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Classificazioni	Descrizione
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.1.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
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

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.11.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.3	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.152.3	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.152.3	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.1	Client a riga di comando Spark.
spark-history-server	2.0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.1.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazioni <code>flink-conf.yaml</code> .
flink-log4j	Modifica le impostazioni <code>Flink log4j.properties</code> .

Classificazioni	Descrizione
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.

Classificazioni	Descrizione
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.

Classificazioni	Descrizione
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.

Classificazioni	Descrizione
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.1.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.1.0. Le modifiche sono relative al rilascio di Amazon EMR 5.0.3.

Data del rilascio: 03 novembre 2016

Modifiche e miglioramenti

- Aggiunto supporto per Flink 1.1.3.
- Presto è stato aggiunto come opzione nella sezione notebook di Hue.

Aggiornamenti

- Aggiornato a HBase 1.2.3
- Aggiornato a Zeppelin 0.6.2

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con le query Tez su Amazon S3 con i file ORC: la query non veniva eseguita correttamente come nelle precedenti versioni di Amazon EMR 4.x.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni

più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion-amzn-EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.11.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
flink-client	1.1.3	Applicazioni e script client a riga di comando Apache Flink.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione

Componente	Versione	Descrizione
		di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.3	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.3	Servizio per servire una o più regioni HBase.
hbase-client	1.2.3	Client di riga di comando HBase.
hbase-rest-server	1.2.3	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.3	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.152.3	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.152.3	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.1	Client a riga di comando Spark.
spark-history-server	2.0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.2	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.1.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
flink-conf	Modifica le impostazione <code>flink-conf.yaml</code> .

Classificazioni	Descrizione
flink-log4j	Modifica le impostazioni Flink log4j.properties.
flink-log4j-yarn-session	Modifica le impostazioni log4j-yarn-session .properties.
flink-log4j-cli	Modifica le impostazioni Flink log4j-cli.properties.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.

Classificazioni	Descrizione
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.0.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
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

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.0.3. Le modifiche sono relative al rilascio di Amazon EMR 5.0.0.

Data del rilascio: 24 ottobre 2016

Aggiornamenti

- Aggiornato a Hadoop 2.7.3
- Aggiornato a Presto 0.152.3, che include il supporto per l'interfaccia Web di Presto. Puoi accedere all'interfaccia Web di Presto sulla porta di utilizzo del coordinatore di Presto 8889. Per ulteriori informazioni sull'interfaccia Web di Presto, consulta [Interfaccia Web](#) nella documentazione di Presto.
- Aggiornato a Spark 2.0.1
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con emr o aws. I

pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.10.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.152.3	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.152.3	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.1	Client a riga di comando Spark.
spark-history-server	2.0.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.0.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.

Classificazioni	Descrizione
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.

Classificazioni	Descrizione
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.

Classificazioni	Descrizione
presto-connector-cassandra	Modifica i valori nel file <code>cassandra.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
presto-connector-jmx	Modifica i valori nel file <code>jmx.properties</code> di Presto.
presto-connector-kafka	Modifica i valori nel file <code>kafka.properties</code> di Presto.
presto-connector-localfile	Modifica i valori nel file <code>localfile.properties</code> di Presto.
presto-connector-mongodb	Modifica i valori nel file <code>mongodb.properties</code> di Presto.
presto-connector-mysql	Modifica i valori nel file <code>mysql.properties</code> di Presto.
presto-connector-postgresql	Modifica i valori nel file <code>postgresql.properties</code> di Presto.
presto-connector-raptor	Modifica i valori nel file <code>raptor.properties</code> di Presto.
presto-connector-redis	Modifica i valori nel file <code>redis.properties</code> di Presto.
presto-connector-tpch	Modifica i valori nel file <code>tpch.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 5.0.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
SDK AWS per Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	-	-	-	-

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.0
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
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

Note di rilascio

Data del rilascio: 27 luglio 2016

Aggiornamenti

- Aggiornato a Hive 2.1
- Aggiornato a Presto 0.150
- Aggiornato a Spark 2.0
- Aggiornato a Hue 3.10.0
- Aggiornato a Pig 0.16.0
- Aggiornato a Tez 0.8.4
- Aggiornato a Zeppelin 0.6.1

Modifiche e miglioramenti

- Amazon EMR supporta le ultime versioni open source di Hive (versione 2.1) e di Pig (versione 0.16.0). Se in passato hai utilizzato Hive o Pig su Amazon EMR, questo potrebbe interessare alcuni casi d'uso. Per ulteriori informazioni, consulta [Hive](#) e [Pig](#).
- Il motore di esecuzione di default per Hive e Pig ora è Tez. Per modificarlo, devi cambiare i valori appropriati nelle classificazioni della configurazione rispettivamente di `hive-site` e `pig-properties`.
- Una caratteristica di debugging della fase potenziata è stata aggiunta per permetterti di vedere la causa sottostante degli errori della fase se il servizio è in grado di determinare la causa. Per ulteriori informazioni, consulta [Debug migliorato delle fasi](#) nella Guida alla gestione di Amazon EMR.
- Le applicazioni che in precedenza terminavano con "-Sandbox" non avranno più quel suffisso. Questo potrebbe spezzare la tua automazione, ad esempio, se stai utilizzando gli script per lanciare i cluster con queste applicazioni. La seguente tabella mostra i nomi dell'applicazione in Amazon EMR 4.7.2 rispetto ad Amazon EMR 5.0.0.

Modifiche al nome dell'applicazione

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandbox	ZooKeeper

- Spark è compilato per Scala 2.11.
- Java 8 è il JVM di default. Tutte le applicazioni vengono eseguite utilizzando il runtime di Java 8. Non ci sono modifiche al target del bytecode dell'applicazione. La maggior parte delle applicazioni continua a scegliere come target Java 7.
- Zeppelin ora include le caratteristiche di autenticazione. Per ulteriori informazioni, consulta [Zeppelin](#).
- Aggiunto supporto per le configurazioni di sicurezza che ti permettono di creare e applicare le opzioni di cifratura più facilmente. Per ulteriori informazioni, consulta [Cifratura dei dati](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.9.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.2-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.7.2-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.2-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	2.1.0-amzn-0	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	2.1.0-amzn-0	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	2.1.0-amzn-0	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	2.1.0-amzn-0	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	2.1.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	2.1.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.10.0-amzn-0	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.46	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.150	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.150	Servizio per l'esecuzione di parti di una query.
pig-client	0.16.0-amzn-0	Client a riga di comando Pig.
spark-client	2.0.0	Client a riga di comando Spark.
spark-history-server	2.0.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	2.0.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	2.0.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.1-SNAPSHOT	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-5.0.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j2	Modifica i valori nel file log4j2.properties i WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-beeline-log4j2	Modifica i valori nel file beeline-log4j2.properties di Hive.

Classificazioni	Descrizione
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j2	Modifica i valori nel file hive-exec-log4j2.properties di Hive.
hive-llap-daemon-log4j2	Modifica i valori nel file llap-daemon-log4j2.properties di Hive.
hive-log4j2	Modifica i valori nel file hive-log4j2.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-hive-site	Modifica i valori nel file hive-site.xml di Spark
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.

Classificazioni	Descrizione
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Versioni del rilascio di Amazon EMR 4.x

Questa sezione contiene le versioni dell'applicazione, le note di rilascio, le versioni dei componenti e le classificazioni di configurazione disponibili in ogni versione di rilascio di Amazon EMR 4.x.

Quando si avvia un cluster, è possibile scegliere tra diversi rilasci di Amazon EMR. In questo modo l'utente può testare e utilizzare le versioni dell'applicazione che soddisfano i propri requisiti di compatibilità. È possibile specificare la versione di rilascio con l'etichetta del rilascio. Le etichette di rilascio sono sotto forma di `emr-x.x.x`. For example, `emr-6.14.0`.

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Per una tabella completa delle versioni delle applicazioni in ogni versione di Amazon EMR 4.x, consulta [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#).

Argomenti

- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)
- [Differenze nelle versioni del rilascio di Amazon EMR 4.x](#)
- [Rilascio 4.9.6 di Amazon EMR](#)
- [Rilascio 4.9.5 di Amazon EMR](#)
- [Rilascio 4.9.4 di Amazon EMR](#)
- [Rilascio 4.9.3 di Amazon EMR](#)
- [Rilascio 4.9.2 di Amazon EMR](#)
- [Rilascio 4.9.1 di Amazon EMR](#)
- [Rilascio 4.8.5 di Amazon EMR](#)
- [Rilascio 4.8.4 di Amazon EMR](#)
- [Rilascio 4.8.3 di Amazon EMR](#)
- [Rilascio 4.8.2 di Amazon EMR](#)
- [Rilascio 4.8.0 di Amazon EMR](#)
- [Rilascio 4.7.4 di Amazon EMR](#)
- [Rilascio 4.7.2 di Amazon EMR](#)
- [Rilascio 4.7.1 di Amazon EMR](#)
- [Rilascio 4.7.0 di Amazon EMR](#)
- [Rilascio 4.6.0 di Amazon EMR](#)
- [Rilascio 4.5.0 di Amazon EMR](#)
- [Rilascio 4.4.0 di Amazon EMR](#)
- [Rilascio 4.3.0 di Amazon EMR](#)
- [Rilascio 4.2.0 di Amazon EMR](#)
- [Rilascio 4.1.0 di Amazon EMR](#)
- [Rilascio 4.0.0 di Amazon EMR](#)

Versioni dell'applicazione nei rilasci di Amazon EMR 4.x

Per una tabella completa che elenca le versioni dell'applicazione disponibili in ogni versione di Amazon EMR 4.x, aprire [Versioni delle applicazioni nelle versioni di Amazon EMR 4.x](#) nel browser.

Differenze nelle versioni del rilascio di Amazon EMR 4.x

La documentazione sulle caratteristiche di Amazon EMR nella Guida alla gestione di Amazon EMR specifica la versione del rilascio di Amazon EMR in cui tale caratteristica è diventata disponibile, nonché differenze applicabili tra caratteristiche di Amazon EMR risalenti alla versione 4.0.0.

A partire dal rilascio di Amazon EMR versione 5.0.0, alcune applicazioni hanno ricevuto un aggiornamento di versione significativo che ha modificato i dettagli di installazione o di funzionamento e altre sono state promosse da applicazioni sandbox ad applicazioni native. Ogni argomento in questa sezione fornisce differenze specifiche dell'applicazione quando si utilizzano le versioni del rilascio di Amazon EMR 4.x.

Argomenti

- [Applicazioni sandbox](#)
- [Considerazioni sull'utilizzo di Hive su Amazon EMR 4.x](#)
- [Considerazioni sull'utilizzo di Pig su Amazon EMR 4.x](#)

Applicazioni sandbox

Quando si utilizzano le versioni del rilascio 4.x di Amazon EMR, alcune applicazioni vengono considerate applicazioni sandbox. Le applicazioni sandbox sono versioni precedenti dell'applicazione che sono state rese disponibili al momento del rilascio di Amazon EMR iniziale a causa della richiesta. Puoi utilizzare la console, AWS CLI o l'API per fare in modo che le applicazioni sandbox vengano installate da Amazon EMR nello stesso modo delle applicazioni native, ma il supporto e la documentazione sono limitati. Le applicazioni sandbox sono diventate applicazioni native, completamente supportate nel rilascio di Amazon EMR versione 5.0.0 e successive. Di seguito sono elencate le applicazioni sandbox nelle versioni del rilascio di Amazon EMR 4.x:

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

Quando installi applicazioni sandbox, i nomi delle applicazioni sono indicati con il suffisso `-sandbox`. Ad esempio, per installare la versione sandbox di *Presto*, utilizza `Presto-sandbox`. Il tempo di

installazione richiesto può essere superiore a quello di un'applicazione completamente supportata. I numeri di versione elencati per ogni applicazione in questa sezione corrispondono alla versione della community dell'applicazione.

Oozie (versioni sandbox)

Oozie è disponibile come un'applicazione sandbox a partire dal rilascio di Amazon EMR versione 4.1.0.

Esempi di Oozie non vengono installati per impostazione predefinita utilizzando le versioni sandbox. Per installare gli esempi, utilizzare SSH per connettersi al nodo master ed eseguire `install-oozie-examples`.

Informazioni sulla versione di Oozie-Sandbox

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
		cemanager, oozie-client, oozie-server
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server
emr-4.9.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
emr-4.9.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
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

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
emr-4.8.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
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-resource-manager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
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-resource-manager, oozie-client, oozie-server
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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
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-httpsfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
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-httpsfs-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-httpsfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Oozie-Sandbox	Componenti installati con Oozie-Sandbox
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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Presto (versioni sandbox)

Presto è disponibile come un'applicazione sandbox a partire dal rilascio di Amazon EMR versione 4.1.0.

Informazioni sulla versione di Presto-Sandbox

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
emr-4.9.6	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
		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, 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

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
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
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

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
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
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

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
emr-4.8.0	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.4	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.2	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

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
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
emr-4.6.0	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
emr-4.5.0	0.140	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
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

Etichetta di rilascio di Amazon EMR	Versione di Presto-Sandbox	Componenti installati con Presto-Sandbox
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
emr-4.1.0	0.119	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Sqoop (versioni sandbox)

Sqoop è disponibile come un'applicazione sandbox a partire dal rilascio di Amazon EMR versione 4.4.0.

Informazioni sulla versione di Sqoop-Sandbox

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
emr-4.9.6	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
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
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

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
emr-4.8.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
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-resource-manager, 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-resource-manager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
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
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

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
emr-4.7.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
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-resource-manager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione di Sqoop-Sandbox	Componenti installati con Sqoop-Sandbox
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-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
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-httplibfs-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-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

Zeppelin (versioni sandbox)

Zeppelin è disponibile come un'applicazione sandbox a partire dal rilascio di Amazon EMR versione 4.1.0.

Informazioni sulla versione di Zeppelin-Sandbox

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.9.6	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
		on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.7.4	0,56	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.2	0,5,6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.7.1	0,5,6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.0	0,5,6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.6.0	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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
emr-4.5.0	0,5,6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.4.0	0,5,6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.3.0	0,5,5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin-Sandbox	Componenti installati con Zeppelin-Sandbox
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 (versioni sandbox)

Zookeeper è disponibile come un'applicazione sandbox a partire dal rilascio di Amazon EMR versione 4.6.0.

Informazioni sulla versione di ZooKeeper-Sandbox

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
emr-4.9.6	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
		hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
emr-4.8.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
emr-4.8.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
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
emr-4.7.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.7.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper-Sandbox	Componenti installati con ZooKeeper-Sandbox
emr-4.6.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Considerazioni sull'utilizzo di Hive su Amazon EMR 4.x

In questa sezione vengono descritte le differenze da considerare quando si utilizza Hive versione 1.0.0 sulle versioni del rilascio 4.x di Amazon EMR, rispetto a Hive 2.x sulle versioni del rilascio 5.x di Amazon EMR.

Transazioni ACID non supportate

Hive su versioni del rilascio 4.x di Amazon EMR non supporta transazioni ACID con dati Hive archiviati in Amazon S3 quando si utilizzano versioni del rilascio 4.x. Se tenti di creare una tabella transazionale in Amazon S3, si verifica un'eccezione.

Lettura e scrittura in tabelle in Amazon S3

Hive su versioni del rilascio 4.x di Amazon EMR è in grado di scrivere direttamente in Amazon S3 senza utilizzare i file temporanei. Ciò consente di migliorare le prestazioni, ma, di conseguenza, non è possibile leggere e scrivere nella stessa tabella in Amazon S3 all'interno della stessa istruzione Hive. Una possibile soluzione è creare e utilizzare una tabella temporanea in HDFS.

Nell'esempio seguente viene mostrato come utilizzare più istruzioni Hive per aggiornare una tabella in Amazon S3. Le istruzioni creano una tabella temporanea in HDFS denominata tmp basata su una tabella in Amazon S3 denominata my_s3_table. La tabella in Amazon S3 viene quindi aggiornata con il contenuto della tabella temporanea.

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;
INSERT OVERWRITE TABLE tmp SELECT ....;
```

```
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

Log4j e Log4j 2

Hive su versioni del rilascio di Amazon EMR 4.x utilizza Log4j. A partire dalla versione 5.0.0, Log4j 2 è l'impostazione predefinita. Queste versioni possono richiedere configurazioni di registrazione diverse. Per dettagli, consulta [Apache Log4j 2](#).

MapReduce è il motore di esecuzione predefinito

Hive su versioni del rilascio di Amazon EMR 4.x utilizza MapReduce come motore di esecuzione predefinito. A partire da Amazon EMR versione 5.0.0, Tez è l'impostazione predefinita, che offre prestazioni migliorate per la maggior parte dei flussi di lavoro.

Autorizzazione Hive

Hive su versioni del rilascio di Amazon EMR 4.x supporta [Autorizzazione Hive](#) per HDFS ma non per EMRFS e Amazon S3. I cluster Amazon EMR vengono eseguiti con l'autorizzazione disabilitata per impostazione predefinita.

Comportamento di unione dei file Hive con Amazon S3

Hive su versioni del rilascio 4.x di Amazon EMR unisce file di piccole dimensioni al termine di un processo di sola mappatura se `hive.merge.mapfiles` è `true`. Un'unione viene attivata solo se la dimensione media dell'output del processo è inferiore all'impostazione `hive.merge.smallfiles.avgsize`. Il comportamento di Amazon EMR Hive è esattamente lo stesso se il percorso di output finale si trova in HDFS. Se il percorso di output si trova in Amazon S3, tuttavia, il parametro `hive.merge.smallfiles.avgsize` viene ignorato. In tal caso, l'attività di unione viene sempre attivata se `hive.merge.mapfiles` è impostato su `true`.

Considerazioni sull'utilizzo di Pig su Amazon EMR 4.x

Pig versione 0.14.0 è installato su cluster creati utilizzando versioni del rilascio di Amazon EMR 4.x. Pig è stato aggiornato alla versione 0.16.0 in Amazon EMR 5.0.0. Di seguito vengono descritte le differenze significative.

Motore di esecuzione predefinito diverso

Pig versione 0.14.0 su versioni del rilascio di Amazon EMR 4.x utilizza MapReduce come motore di esecuzione predefinito. Pig versione 0.16.0 e successive utilizzano Apache Tez. Puoi impostare in

maniera esplicita `execType=mapreduce` nella classificazione di configurazione `pig-properties` per utilizzare MapReduce.

Funzioni Pig definite dall'utente (UDF) interrotte

Le UDF personalizzate che erano disponibili in Pig su versioni del rilascio 4.x di Amazon EMR sono state interrotte a partire da Pig 0.16.0. La maggior parte delle UDF dispongono di funzioni equivalenti che è possibile utilizzare. Nella tabella seguente vengono elencate le FDU interrotte e le funzioni equivalenti. Per ulteriori informazioni, consulta [Built-in functions \(Funzioni integrate\)](#) sul sito Apache Pig.

FDU interrotta	Funzione equivalente
<code>FORMAT_DT(dtformat, date)</code>	<code>GetHour(date)</code> , <code>GetMinute(date)</code> , <code>GetMonth(date)</code> , <code>GetSecond(date)</code> , <code>GetWeek(date)</code> , <code>GetYear(date)</code> , <code>GetDay(date)</code>
<code>EXTRACT(string, pattern)</code>	<code>REGEX_EXTRACT_ALL(string, pattern)</code>
<code>REPLACE(string, pattern, replacement)</code>	<code>REPLACE(string, pattern, replacement)</code>
<code>DATE_TIME()</code>	<code>ToDate()</code>
<code>DURATION(dt, dt2)</code>	<code>WeeksBetween(dt, dt2)</code> , <code>YearsBetween(dt, dt2)</code> , <code>SecondsBetween(dt, dt2)</code> , <code>MonthsBetween(dt, dt2)</code> , <code>MinutesBetween(dt, dt2)</code> , <code>HoursBetween(dt, dt2)</code>
<code>EXTRACT_DT(format, date)</code>	<code>GetHour(date)</code> , <code>GetMinute(date)</code> , <code>GetMonth(date)</code> , <code>GetSecond(date)</code> , <code>GetWeek(date)</code> , <code>GetYear(date)</code> , <code>GetDay(date)</code>
<code>OFFSET_DT(date, duration)</code>	<code>AddDuration(date, duration)</code> , <code>SubtractDuration(date, duration)</code>
<code>PERIOD(dt, dt2)</code>	<code>WeeksBetween(dt, dt2)</code> , <code>YearsBetween(dt, dt2)</code> , <code>SecondsBetween(dt, dt2)</code> , <code>MonthsBetween(dt, dt2)</code> , <code>MinutesBetween(dt, dt2)</code> , <code>HoursBetween(dt, dt2)</code>

FDU interrotta	Funzione equivalente
CAPITALIZE(string)	UCFIRST(string)
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

Le seguenti FDU sono state interrotte senza un equivalente: FORMAT(), LOCAL_DATE(), LOCAL_TIME(), CENTER(), LEFT_PAD(), REPEAT(), REPLACE_ONCE(), RIGHT_PAD(), STRIP(), STRIP_END(), STRIP_START(), SWAP_CASE().

Comandi Grunt non più disponibili

Alcuni comandi Grunt non sono più disponibili a partire da Pig 0.16.0. Nella tabella seguente vengono elencati i comandi Grunt in Pig 0.14.0 e i comandi equivalenti nella versione corrente, dove applicabile.

Pig 0.14.0 e comandi Grunt correnti equivalenti

Comando Grunt Pig 0.14.0	Comando Grunt in Pig 0.16.0 e versioni successive
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	Nessun equivalente
ls <non-hdfs-path>;	fs -ls <non-hdfs-path>;
move <non-hdfs-path> <non-hdfs-path>;	fs -mv <non-hdfs-path> <non-hdfs-path>;
copy <non-hdfs-path> <non-hdfs-path>;	fs -cp <non-hdfs-path> <non-hdfs-path>;
copyToLocal <non-hdfs-path> <local-path>;	fs -copyToLocal <non-hdfs-path> <local-path>;

Comando Grunt Pig 0.14.0	Comando Grunt in Pig 0.16.0 e versioni successive
<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>

Funzionalità rimosse per home directory non HDFS

Pig 0.14.0 su versioni del rilascio 4.x di Amazon EMR dispone di due meccanismi per consentire a utenti diversi dall'utente `hadoop`, che non dispongono di home directory, di eseguire script Pig. Il primo meccanismo è un fallback automatico che consente di impostare la directory di lavoro iniziale sulla directory principale se la home directory non esiste. Il secondo è una proprietà `pig.initial.fs.name` che consente di modificare la directory di lavoro iniziale.

Questi meccanismi non sono disponibili a partire da Amazon EMR versione 5.0.0 e gli utenti devono disporre di una home directory su HDFS. Ciò non si applica all'utente `hadoop`, perché una home directory viene fornita all'avvio. Gli script eseguiti utilizzando fasi `jar Hadoop` sono impostati per default sull'utente `Hadoop`, a meno che non venga specificato in modo esplicito un altro utente utilizzando `command-runner.jar`.

Rilascio 4.9.6 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,73
Hive	1.0.0	1.0.0	1.0.0	1.0.0

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
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
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-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
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

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
<code>ganglia-web</code>	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.9.6

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.

Classificazioni	Descrizione
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.

Classificazioni	Descrizione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.9.5 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,73
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-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 4.9.5. Le modifiche sono relative alla versione 4.9.4.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- HBase
 - Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.

Componente	Versione	Descrizione
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.

Componente	Versione	Descrizione
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.9.5

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.

Classificazioni	Descrizione
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.

Classificazioni	Descrizione
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.9.4 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Hadoop	2,73	2,73	2,73	2,73
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
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
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

Note di rilascio

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 4.9.4. Le modifiche sono relative alla versione 4.9.3.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri

Componente	Versione	Descrizione
		raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni di emr-4.9.4

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.

Classificazioni	Descrizione
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.9.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,73
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	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.3. Le modifiche sono relative al rilascio di Amazon EMR 4.9.2.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni di emr-4.9.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.

Classificazioni	Descrizione
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.

Classificazioni	Descrizione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.9.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,73
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-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.2. Le modifiche sono relative al rilascio di Amazon EMR 4.9.1.

Data del rilascio: 13 luglio 2017

Modifiche minori, correzioni dei bug e miglioramenti sono stati effettuati in questo rilascio.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.3.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emrfs	2.17.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.

Componente	Versione	Descrizione
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni di emr-4.9.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.

Classificazioni	Descrizione
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.

Classificazioni	Descrizione
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.9.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Hadoop	2,73	2,73	2,73	2,73
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
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.1. Le modifiche sono relative al rilascio di Amazon EMR 4.8.4.

Data del rilascio: 10 aprile 2017

Problemi noti risolti dai rilasci precedenti

- Backport di [HIVE-9976](#) e [HIVE-10106](#)
- Risolto un problema in YARN in cui un numero elevato di nodi (superiore a 2.000) e di container (superiore a 5.000) causa un errore di esaurimento della memoria, ad esempio: "Exception in thread main java.lang.OutOfMemoryError".

Modifiche e miglioramenti

- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2017.03. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- È stato rimosso Python 2.6 dall'immagine Linux di base Amazon EMR. Puoi installare Python 2.6 manualmente se necessario.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.3.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.15.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.3-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-9	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	1.0.0-amzn-9	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-9	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-9	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-9	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-9	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.9.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.

Classificazioni	Descrizione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.

Classificazioni	Descrizione
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.8.5 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,73
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
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
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.8

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.14.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-ftpfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-8	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-8	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-8	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-8	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	1.0.0-amzn-8	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-8	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.8.5

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.

Classificazioni	Descrizione
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.

Classificazioni	Descrizione
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.

Classificazioni	Descrizione
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.8.4 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.0
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,73	2,73	2,73	2,72
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-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.0
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
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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.4. Le modifiche sono relative al rilascio di Amazon EMR 4.8.3.

Data di rilascio: 7 febbraio 2017

Modifiche minori, correzioni dei bug e miglioramenti sono stati effettuati in questo rilascio.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.2.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.

Componente	Versione	Descrizione
emrfs	2.14.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.

Componente	Versione	Descrizione
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.

Componente	Versione	Descrizione
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-8	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-8	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-8	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-8	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-8	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-8	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.

Componente	Versione	Descrizione
mysql-server	5.5.54+	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.

Componente	Versione	Descrizione
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.25+	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.8.4

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.

Classificazioni	Descrizione
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.

Classificazioni	Descrizione
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.

Classificazioni	Descrizione
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.8.3 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)

- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.8.3	emr-4.8.2	emr-4.8.0	emr-4.7.4
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-4.8.3	emr-4.8.2	emr-4.8.0	emr-4.7.4
Hadoop	2,73	2,73	2,72	2,72
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	0.148
Spark	1.6.3	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.0	emr-4.7.4
Tez	0.8.4	0.8.4	0.8.4	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0,5,6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.3. Le modifiche sono relative al rilascio di Amazon EMR 4.8.2.

Data del rilascio: 29 dicembre 2016

Aggiornamenti

- Aggiornato a Presto 0.157.1. Per ulteriori informazioni, consulta la sezione delle [note di rilascio di Presto](#) nella documentazione di Presto.
- Aggiornato a Spark 1.6.3. Per ulteriori informazioni, consulta [Note di rilascio di Spark](#) nella documentazione di Apache Spark.
- Aggiornato a Zookeeper 3.4.9. Per ulteriori informazioni, consulta [Note di rilascio di ZooKeeper](#) nella documentazione di Apache ZooKeeper.

Modifiche e miglioramenti

- Aggiunto supporto per il tipo di istanza Amazon EC2 m4.16xlarge in Amazon EMR versione 4.8.3 e successive, escluse 5.0.0, 5.0.3 e 5.2.0.

- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in Hadoop in cui il thread ReplicationMonitor poteva bloccarsi per molto tempo a causa di un conflitto tra la replica e la cancellazione dello stesso file in un cluster di grandi dimensioni.
- Risolto un problema in cui ControlledJob#toString restituiva un errore con una null pointer exception (NPE) quando lo stato del processo non era aggiornato correttamente.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	4.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.2.0	Librerie utili per l'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.13.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-1	Libreria e client di riga di comando HDFS

Componente	Versione	Descrizione
hadoop-hdfs-namenode	2.7.3-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.3-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.3-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.

Componente	Versione	Descrizione
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-8	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-8	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-8	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-8	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-8	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-8	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accessi in formato JSON all'API Avatica
presto-coordinator	0.157.1	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.157.1	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.3	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	1.6.3	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.3	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.3	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.9	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.9	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.8.3

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file <code>hbase-log4j.properties</code> di HBase.
hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.

Classificazioni	Descrizione
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.

Classificazioni	Descrizione
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.

Classificazioni	Descrizione
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.8.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.8.2	emr-4.8.0	emr-4.7.4	emr-4.7.2
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.2	emr-4.8.0	emr-4.7.4	emr-4.7.2
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	2,73	2,72	2,72	2,72
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	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2

	emr-4.8.2	emr-4.8.0	emr-4.7.4	emr-4.7.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	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

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.2. Le modifiche sono relative al rilascio di Amazon EMR 4.8.0.

Data del rilascio: 24 ottobre 2016

Aggiornamenti

- Aggiornato a Hadoop 2.7.3
- Aggiornato a Presto 0.152.3, che include il supporto per l'interfaccia Web di Presto. Puoi accedere all'interfaccia Web di Presto sulla porta di utilizzo del coordinatore di Presto 8889. Per ulteriori informazioni sull'interfaccia Web di Presto, consulta [Interfaccia Web](#) nella documentazione di Presto.
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	4.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.1.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.10.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.3-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.3-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.3-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.3-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.3-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.3-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.3-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.3-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-7	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	1.0.0-amzn-7	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-7	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-7	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-7	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-7	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.52	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.152.3	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.152.3	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.2	Client a riga di comando Spark.
spark-history-server	1.6.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.8.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.

Classificazioni	Descrizione
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.

Classificazioni	Descrizione
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.

Classificazioni	Descrizione
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.8.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.8.0	emr-4.7.4	emr-4.7.2	emr-4.7.1
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2,72	2,72	2,72	2,72
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.8.0	emr-4.7.4	emr-4.7.2	emr-4.7.1
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	0.151	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	0.8.4	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

	emr-4.8.0	emr-4.7.4	emr-4.7.2	emr-4.7.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.0. Le modifiche sono relative al rilascio di Amazon EMR 4.7.2.

Data del rilascio: 7 settembre 2016

Aggiornamenti

- Aggiornato a HBase 1.2.2
- Aggiornato a Presto-Sandbox 0.151
- Aggiornato a Tez 0.8.4
- Aggiornato a Zeppelin-Sandbox 0.6.1

Modifiche e miglioramenti

- Risolto un problema con YARN dove ApplicationMaster cercava di pulire i container che non esistono più perché le loro istanze sono state terminate.
- Corrette le azioni hive-server2 URL per Hive2 negli esempi Oozie.
- Aggiunto supporto per i cataloghi Presto aggiuntivi.
- Backport per patch: [HIVE-8948](#), [HIVE-12679](#), [HIVE-13405](#), [PHOENIX-3116](#), [HADOOP-12689](#)
- Aggiunto supporto per le configurazioni di sicurezza che ti permettono di creare e applicare le opzioni di cifratura più facilmente. Per ulteriori informazioni, consulta [Cifratura dei dati](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con emr o aws. I

pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	3.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.9.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.2-amzn-4	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-4	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-4	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-4	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.2-amzn-4	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-4	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-4	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-4	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.7.2-amzn-4	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.2	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.2	Servizio per servire una o più regioni HBase.
hbase-client	1.2.2	Client di riga di comando HBase.
hbase-rest-server	1.2.2	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.2	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-7	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-7	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	1.0.0-amzn-7	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-7	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-7	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-7	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.51	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client

Componente	Versione	Descrizione
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.151	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.151	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.2	Client a riga di comando Spark.
spark-history-server	1.6.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.4	L'applicazione e le librerie tez YARN.

Componente	Versione	Descrizione
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.6.1	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.8.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.

Classificazioni	Descrizione
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hiveserver2-site	Modifica i valori nel file hiveserver2-site.xml di Hive Server2
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-blackhole	Modifica i valori nel file blackhole.properties di Presto.
presto-connector-cassandra	Modifica i valori nel file cassandra.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
presto-connector-jmx	Modifica i valori nel file jmx.properties di Presto.

Classificazioni	Descrizione
presto-connector-kafka	Modifica i valori nel file kafka.properties di Presto.
presto-connector-localfile	Modifica i valori nel file localfile.properties di Presto.
presto-connector-mongodb	Modifica i valori nel file mongodb.properties di Presto.
presto-connector-mysql	Modifica i valori nel file mysql.properties di Presto.
presto-connector-postgresql	Modifica i valori nel file postgresql.properties di Presto.
presto-connector-raptor	Modifica i valori nel file raptor.properties di Presto.
presto-connector-redis	Modifica i valori nel file redis.properties di Presto.
presto-connector-tpch	Modifica i valori nel file tpch.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.7.4 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.7.4	emr-4.7.2	emr-4.7.1	emr-4.7.0
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,75
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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	2,72	2,72	2,72	2,72
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	-	-	-	-

	emr-4.7.4	emr-4.7.2	emr-4.7.1	emr-4.7.0
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.0	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.147	0.147
Spark	1.6.2	1.6.2	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
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	-	-	-	-

	emr-4.7.4	emr-4.7.2	emr-4.7.1	emr-4.7.0
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

Note di rilascio

Questa è una versione di patch da aggiungere l'autenticazione AWS Signature Version 4 per le richieste inviate ad Amazon S3. Tutte le applicazioni e i componenti sono gli stessi del rilascio precedente di Amazon EMR.

Important

In questa versione, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste ad Amazon S3. Per ulteriori informazioni, consulta [Novità](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	3.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.1.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.8.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.2-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.2-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.2-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.

Componente	Versione	Descrizione
hbase-hmaster	1.2.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.1	Servizio per servire una o più regioni HBase.
hbase-client	1.2.1	Client di riga di comando HBase.
hbase-rest-server	1.2.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-6	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-6	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-6	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-6	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	1.0.0-amzn-6	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-6	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.46	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.148	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.148	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.2	Client a riga di comando Spark.
spark-history-server	1.6.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.3	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni di emr-4.7.4

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop

Classificazioni	Descrizione
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.

Classificazioni	Descrizione
phoenix-hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di Phoenix.
phoenix-hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di Phoenix.
phoenix-log4j	Modifica i valori nel file <code>log4j.properties</code> di Phoenix.
phoenix-metrics	Modifica i valori nel file <code>hadoop-metrics2-phoenix.properties</code> di Phoenix.
pig-properties	Modifica i valori nel file <code>pig.properties</code> di Pig.
pig-log4j	Modifica i valori nel file <code>log4j.properties</code> di Pig.
presto-log	Modifica i valori nel file <code>log.properties</code> di Presto.
presto-config	Modifica i valori nel file <code>config.properties</code> di Presto.
presto-connector-hive	Modifica i valori nel file <code>hive.properties</code> di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file <code>spark-defaults.conf</code> di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file <code>log4j.properties</code> di Spark.
spark-metrics	Modifica i valori nel file <code>metrics.properties</code> di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.

Classificazioni	Descrizione
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.7.2 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.0
SDK AWS per Java	1,10,75	1,10,75	1,10,75	1,10,27
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2,72	2,72	2,72	2,72
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	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.0
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.147	0.147	0.143
Spark	1.6.2	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
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	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.0
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

Note di rilascio

Le seguenti note di rilascio includono informazioni su Amazon EMR 4.7.2.

Data del rilascio: 15 luglio 2016

Funzionalità

- Aggiornato a Mahout 0.12.2
- Aggiornato a Presto 0.148
- Aggiornato a Spark 1.6.2
- Ora puoi creare un `AWSCredentialsProvider` da utilizzare con EMRFS utilizzando un URI come parametro. Per ulteriori informazioni, consulta [Crea un AWSCredentialsProvider per EMRFS](#).
- EMRFS ora consente agli utenti di configurare un endpoint DynamoDB personalizzato per i metadata Consistent View (Visualizzazione coerente) utilizzando la proprietà `fs.s3.consistent.dynamodb.endpoint` in `emrfs-site.xml`.
- Aggiunto uno script in `/usr/bin` chiamato `spark-example`, che esegue il wrapping di `/usr/lib/spark/spark/bin/run-example` in modo che tu possa eseguire gli esempi direttamente. Ad esempio, per eseguire un esempio SparkPi disponibile con la distribuzione Spark, puoi eseguire `spark-example SparkPi 100` dalla riga di comando o utilizzando `command-runner.jar` come fase nella API.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui Oozie aveva `spark-assembly.jar` in una posizione non corretta quando Spark era installato, il che portava a un errore durante il lancio delle applicazioni Spark con Oozie.
- Risolto un problema con il logging Spark Log4j-based nei container YARN.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.2.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.1.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.8.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.2-amzn-3	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-3	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-3	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-3	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.2-amzn-3	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-3	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-3	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.2-amzn-3	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.1	Servizio per servire una o più regioni HBase.
hbase-client	1.2.1	Client di riga di comando HBase.
hbase-rest-server	1.2.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-6	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	1.0.0-amzn-6	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-6	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-6	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-6	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-6	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.2	Librerie per machine learning.
mysql-server	5.5.46	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica
presto-coordinator	0.148	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.148	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.2	Client a riga di comando Spark.
spark-history-server	1.6.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.2	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.

Componente	Versione	Descrizione
tez-on-yarn	0.8.3	L'applicazione e le librerie tez YARN.
webserver	2.4.23	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni di emr-4.7.2

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.

Classificazioni	Descrizione
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hadoop-ssl-server	Modifica la configurazione server ssl hadoop
hadoop-ssl-client	Modifica la configurazione client ssl hadoop
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.7.1 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.7.1	emr-4.7.0	emr-4.6.0	emr-4.5.0
SDK AWS per Java	1,10,75	1,10,75	1,10,27	1,10,27
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2,72	2,72	2,72	2,72
Hive	1.0.0	1.0.0	1.0.0	1.0.0

	emr-4.7.1	emr-4.7.0	emr-4.6.0	emr-4.5.0
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	0.140
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	0.8.3	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.0	emr-4.5.0
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	3.4.8	-

Note di rilascio

Le seguenti note di rilascio includono informazioni su Amazon EMR 4.7.1.

Data del rilascio: 10 giugno 2016

Problemi noti risolti dai rilasci precedenti

- Risolto un problema che prolungava i tempi di avvio dei cluster lanciati in un VPC con sottoreti private. Il bug interessava soltanto i cluster lanciati con il rilascio di Amazon EMR 4.7.0.
- Risolto un problema che gestiva in modo non corretto l'elenco dei file in Amazon EMR per i cluster lanciati con il rilascio di Amazon EMR 4.7.0.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`.

EmrVersion inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.0.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.7.1	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
<code>ganglia-web</code>	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.2-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.2-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.2-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.1	Servizio per servire una o più regioni HBase.
hbase-client	1.2.1	Client di riga di comando HBase.
hbase-rest-server	1.2.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-5	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-5	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-5	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	1.0.0-amzn-5	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-5	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-5	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.0	Librerie per Machine Learning.
mysql-server	5.5.46	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.147	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.147	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.1	Client a riga di comando Spark.
spark-history-server	1.6.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.3	L'applicazione e le librerie tez YARN.
webserver	2.4.18	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.7.1

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file <code>hbase-log4j.properties</code> di HBase.

Classificazioni	Descrizione
hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file <code>log4j.properties</code> di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file <code>hive-exec-log4j.properties</code> di Hive.
hive-log4j	Modifica i valori nel file <code>hive-log4j.properties</code> di Hive.

Classificazioni	Descrizione
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.7.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.7.0	emr-4.6.0	emr-4.5.0	emr-4.4.0
SDK AWS per Java	1,10,75	1,10,27	1,10,27	1,10,27
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.0	-	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2,72	2,72	2,72	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	0.12.0	0.11.1	0.11.1	0.11.1

	emr-4.7.0	emr-4.6.0	emr-4.5.0	emr-4.4.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.143	0.140	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	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	-	-

Note di rilascio

Important

Amazon EMR 4.7.0 è una versione obsoleta. Utilizza invece Amazon EMR 4.7.1 o versioni successive.

Data del rilascio: 2 giugno 2016

Funzionalità

- Aggiunto Apache Phoenix 4.7.0
- Aggiunto Apache Tez 0.8.3
- Aggiornato a HBase 1.2.1
- Aggiornato a Mahout 0.12.0
- Aggiornato a Presto 0.147
- Aggiornato AWS SDK for Java a 1.10.75
- Il flag finale è stato rimosso dalla proprietà `mapreduce.cluster.local.dir` in `mapred-site.xml` per permettere agli utenti di eseguire Pig in modalità locale.
- Driver Amazon Redshift JDBC disponibili sul cluster

I driver JDBC Amazon Redshift sono ora inclusi in `/usr/share/aws/redshift/jdbc./usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` è il driver JDBC Amazon Redshift compatibile con 4.1 e `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` è il driver JDBC Amazon Redshift compatibile con 4.0. Per ulteriori informazioni, vedere [Configurazione di una connessione JDBC](#) nella Guida alla gestione di Amazon Redshift.

- Java 8

A eccezione di Presto, OpenJDK 1.7 è il JDK di default utilizzato per tutte le applicazioni. Tuttavia, sia OpenJDK 1.7 sia OpenJDK 1.8 sono installati. Per informazioni su come impostare `JAVA_HOME` per le applicazioni, consulta [Configurare le applicazioni per usare Java 8](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con il kernel che influiva in modo significativo sulla performance dei volumi Throughput Optimized HDD (st1) EBS per Amazon EMR in emr-4.6.0.

- Risolto un problema in cui un cluster restituiva un errore se una zona di cifratura HDFS veniva specificata senza scegliere Hadoop come applicazione.
- Modificata la policy di scrittura HDFS di default da RoundRobin a AvailableSpaceVolumeChoosingPolicy. Alcuni volumi non venivano utilizzati correttamente con la configurazione RoundRobin, il che portava i nodi principali a restituire un errore e alla creazione di HDFS non affidabili.
- Risolto un problema con la CLI EMRFS che causava un'eccezione durante la creazione della tabella dei metadati DynamoDB di default per viste uniformi.
- Risolto un problema di deadlock in EMRFS che accadeva eventualmente durante le operazioni multipartite di rinomina e copia.
- Risolto un problema con EMRFS che faceva sì che la dimensione di default di CopyPart fosse 5 MB. Il valore di default è ora di 128 MB.
- Risolto un problema con la configurazione upstart di Zeppelin che eventualmente ti impediva di poter arrestare il servizio.
- Risolto un problema con Spark e Zeppelin, che ti impediva di utilizzare lo schema URI `s3a://` perché `/usr/lib/hadoop/hadoop-aws.jar` non era caricato correttamente nel rispettivo classpath.
- Backport per [HUE-2484](#).
- Backport per [commit](#) da Hue 3.9.0 (nessun JIRA esistente) per risolvere un problema con il campione del browser HBase.
- Backport per [HIVE-9073](#).

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato

myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	3.1.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.2.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.4.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.7.1	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.

Componente	Versione	Descrizione
hadoop-client	2.7.2-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.7.2-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.

Componente	Versione	Descrizione
hadoop-yarn-timeline-server	2.7.2-amzn-2	Servizio per il recupero di informazioni correnti e della cronologia per applicazioni YARN.
hbase-hmaster	1.2.1	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.1	Servizio per servire una o più regioni HBase.
hbase-client	1.2.1	Client di riga di comando HBase.
hbase-rest-server	1.2.1	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.1	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-5	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-5	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-5	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.

Componente	Versione	Descrizione
hive-client	1.0.0-amzn-5	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-5	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-5	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-7	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.12.0	Librerie per Machine Learning.
mysql-server	5.5.46	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
phoenix-library	4.7.0-HBase-1.2	Le librerie Phoenix per server e client
phoenix-query-server	4.7.0-HBase-1.2	Un server leggero che fornisce accesso JDBC nonché buffer di protocollo e accesso in formato JSON all'API Avatica

Componente	Versione	Descrizione
presto-coordinator	0.147	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.147	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.1	Client a riga di comando Spark.
spark-history-server	1.6.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
tez-on-yarn	0.8.3	L'applicazione e le librerie tez YARN.
webserver	2.4.18	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.

Componente	Versione	Descrizione
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.7.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file <code>hbase-log4j.properties</code> di HBase.

Classificazioni	Descrizione
hbase-metrics	Modifica i valori nel file <code>hadoop-metrics2-hbase.properties</code> di HBase.
hbase-policy	Modifica i valori nel file <code>hbase-policy.xml</code> di HBase.
hbase-site	Modifica i valori nel file <code>hbase-site.xml</code> di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file <code>log4j.properties</code> di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file <code>webhcat-site.xml</code> file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file <code>hive-exec-log4j.properties</code> di Hive.
hive-log4j	Modifica i valori nel file <code>hive-log4j.properties</code> di Hive.

Classificazioni	Descrizione
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
phoenix-hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di Phoenix.
phoenix-hbase-site	Modifica i valori nel file hbase-site.xml di Phoenix.

Classificazioni	Descrizione
phoenix-log4j	Modifica i valori nel file log4j.properties di Phoenix.
phoenix-metrics	Modifica i valori nel file hadoop-metrics2-phoenix.properties di Phoenix.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
tez-site	Modifica i valori nel file tez-site.xml file di Tez.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.6.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
SDK AWS per Java	1,10,27	1,10,27	1,10,27	1,10,27
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2,72	2,72	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

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.140	0.136	0.130
Spark	1.6.1	1.6.1	1.6.0	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,6	0,5,6	0,5,6	0,5,5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.6.0.

- Aggiunto HBase 1.2.0
- Aggiunto Zookeeper-Sandbox 3.4.8
- Aggiornato a Presto-Sandbox 0.143
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.03.0. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.
- Problema che riguarda i tipi di volume Throughput Optimized HDD (st1) EBS

Un problema nel kernel Linux versioni 4.2 e superiori influisce in modo significativo sulla performance dei volumi Throughput Optimized HDD (st1) EBS per EMR. Questo rilascio (emr-4.6.0) utilizza kernel versione 4.4.5 per cui è interessato dal problema. Per questo motivo ti consigliamo di non usare emr-4.6.0 se desideri usare i volumi st1 EBS. Puoi utilizzare i rilasci emr-4.5.0 o Amazon EMR precedenti con st1 senza che si crei alcun problema. Inoltre, forniamo la correzione per i rilasci futuri.

- Valori di default di Python

Python 3.4 ora è installato di default, ma Python 2.7 rimane la versione di default del sistema. Puoi configurare Python 3.4 come la versione di default del sistema utilizzando un'operazione di bootstrap; puoi anche usare la configurazione API per impostare l'export `PYSPARK_PYTHON` su `/usr/bin/python3.4` nella classificazione `spark-env` in modo che interessi la versione di Python utilizzata da PySpark.

- Java 8

A eccezione di Presto, OpenJDK 1.7 è il JDK di default utilizzato per tutte le applicazioni. Tuttavia, sia OpenJDK 1.7 sia OpenJDK 1.8 sono installati. Per informazioni su come impostare `JAVA_HOME` per le applicazioni, consulta [Configurare le applicazioni per usare Java 8](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui il provisioning dell'applicazione poteva restituire un errore in modo casuale a causa della password generata.
- In precedenza, `mysqld` era installato su tutti i nodi. Ora, è installato soltanto sull'istanza master e solo se l'applicazione scelta include `mysql-server` come componente. Attualmente, le seguenti applicazioni includono il componente `mysql-server`: HCatalog, Hive, Hue, Presto-Sandbox e Sqoop-Sandbox.
- Modificato `yarn.scheduler.maximum-allocation-vcores` a 80 rispetto al valore di default di 32. Questa scelta risolve un errore introdotto in emr-4.4.0 che si verificava principalmente

con Spark durante l'uso dell'opzione `maximizeResourceAllocation` in un cluster il cui tipo di istanza core era uno dei pochi tipi di istanza di grandi dimensioni che hanno i vcore YARN impostati su un valore superiore a 32; tra questi `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` o `m4.10xlarge` erano interessati dal problema.

- `s3-dist-cp` ora utilizza EMRFS per tutte le nomine Amazon S3 e non viene più gestito su una directory HDFS temporanea.
- Risolto un problema con la gestione dell'eccezione per il caricamento in più parti della crittografia lato client.
- Aggiunta un'opzione per permettere agli utenti di modificare la classe di archiviazione Amazon S3. Per default, l'impostazione è `STANDARD`. L'impostazione della classificazione della configurazione `emrfs-site` è `fs.s3.storageClass` e i valori possibili sono `STANDARD`, `STANDARD_IA` e `REDUCED_REDUNDANCY`. Per ulteriori informazioni sulle classi di archiviazione, consulta [Classi di archiviazione](#) nella Guida per l'utente di Amazon Simple Storage Service.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.3.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.6.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.2-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.2-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.2-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.2-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hbase-hmaster	1.2.0	Servizio per un cluster HBase responsabile del coordinamento delle regioni e dell'esecuzione di comandi amministrativi.
hbase-region-server	1.2.0	Servizio per servire una o più regioni HBase.

Componente	Versione	Descrizione
hbase-client	1.2.0	Client di riga di comando HBase.
hbase-rest-server	1.2.0	Servizio che fornisce un endpoint RESTful HTTP per HBase.
hbase-thrift-server	1.2.0	Servizio che fornisce un endpoint Thrift a HBase.
hcatalog-client	1.0.0-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-4	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-4	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.

Componente	Versione	Descrizione
hue-server	3.7.1-amzn-6	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.1	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
presto-coordinator	0.143	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.143	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.1	Client a riga di comando Spark.
spark-history-server	1.6.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.1	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	1.6.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
webserver	2.4	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.
zookeeper-server	3.4.8	Servizio centralizzato per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo.
zookeeper-client	3.4.8	Client a riga di comando ZooKeeper.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.6.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.

Classificazioni	Descrizione
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hbase-env	Modifica i valori nell'ambiente HBase.
hbase-log4j	Modifica i valori nel file hbase-log4j.properties di HBase.
hbase-metrics	Modifica i valori nel file hadoop-metrics2-hbase.properties di HBase.
hbase-policy	Modifica i valori nel file hbase-policy.xml di HBase.
hbase-site	Modifica i valori nel file hbase-site.xml di HBase.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.

Classificazioni	Descrizione
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.
zookeeper-config	Modifica i valori nel file zoo.cfg di ZooKeeper.
zookeeper-log4j	Modifica i valori nel file log4j.properties di ZooKeeper.

Rilascio 4.5.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) e [Zeppelin-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
SDK AWS per Java	1,10,27	1,10,27	1,10,27	1,10,27
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-
HCatalog	1.0.0	1.0.0	-	-
Hadoop	2,72	2.7.1	2.7.1	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.0	0.11.0

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
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	-	-	-	-

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.5.0.

Data del rilascio: 4 aprile 2016

Funzionalità

- Aggiornato a Spark 1.6.1
- Aggiornato a Hadoop 2.7.2
- Aggiornato a Presto 0.140
- Aggiunto supporto AWS KMS per la crittografia lato server Amazon S3.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui i server MySQL e Apache non si riavviavano dopo il riavvio di un nodo.
- Risolto un problema in cui IMPORT non funzionava correttamente con le tabelle non partizionate archiviate in Amazon S3
- Risolto un problema con Presto dove si richiedeva che la directory di gestione temporanea fosse /mnt/tmp anziché /tmp quando si scrive sulla tabelle Hive.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.2.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.5.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.2-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".

Componente	Versione	Descrizione
hadoop-hdfs-datanode	2.7.2-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.2-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.2-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.2-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.2-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.2-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.2-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hcatalog-client	1.0.0-amzn-4	Client a riga di comando "hcat" per la manipolazione del server hcatalog.

Componente	Versione	Descrizione
hcatalog-server	1.0.0-amzn-4	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.
hcatalog-webhcat-server	1.0.0-amzn-4	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-4	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-4	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-4	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-5	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.1	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
presto-coordinator	0.140	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.140	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.1	Client a riga di comando Spark.
spark-history-server	1.6.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.1	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.1	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
webserver	2.4	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.5.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file <code>jndi.properties</code> di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file <code>proto-hive-site.xml</code> di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file <code>log4j.properties</code> di WebHCat HCatalog.

Classificazioni	Descrizione
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.

Classificazioni	Descrizione
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
yarn-env	Modifica i valori nell'ambiente YARN.

Classificazioni	Descrizione
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Rilascio 4.4.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) e [Zeppelin-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
SDK AWS per Java	1,10,27	1,10,27	1,10,27	Non disponibile
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
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

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
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-SNA PSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.4.0.

Data di rilascio: 14 marzo 2016

Funzionalità

- Aggiunto HCatalog 1.0.0
- Aggiunto Sqoop-Sandbox 1.4.6
- Aggiornato a Presto 0.136

- Aggiornato a Zeppelin 0.5.6
- Aggiornato a Mahout 0.11.1
- Abilitato `dynamicResourceAllocation` per impostazione predefinita.
- Aggiunta una tabella di tutte le classificazioni della configurazione per il rilascio. Per ulteriori informazioni, consulta la tabella sulle classificazioni della configurazione in [Configurazione delle applicazioni](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui l'impostazione `maximizeResourceAllocation` non avrebbe riservato memoria sufficiente per i daemon YARN ApplicationMaster.
- Risolto un problema rilevato con un DNS personalizzato. Se le voci in `resolve.conf` precedono le voci personalizzate fornite, allora le voci personalizzate non sono risolvibili. Questo comportamento è stato influenzato dai cluster in un VPC dove il server di nomi VPC di default era inserito come prima voce in `resolve.conf`.
- Risolto un problema in cui il Python di default era spostato nella versione 2.7 e boto non era installato per quella versione.
- Risolto un problema in cui i container YARN e le applicazioni Spark generavano un file unico del round robin database (rrd) di Ganglia il che portava al collegamento del primo disco al riempimento dell'istanza. Per via di questa correzione, i parametri a livello del container YARN sono stati disattivati così come quelli a livello dell'applicazione Spark.
- Risolto un problema in log pusher che avrebbe cancellato tutte le cartelle vuote dei log. La conseguenza era che Hive CLI non era in grado di effettuare l'accesso perché log pusher cancellava la cartella vuota `user` sotto `/var/log/hive`.
- Risolto un problema che riguardava le importazioni di Hive, che interessava il partizionamento e produceva un errore durante l'importazione.
- Risolto un problema in cui EMRFS e `s3-dist-cp` non gestivano correttamente i nomi dei bucket che contengono punti.
- Modificato un comportamento in EMRFS che prevedeva che nei bucket con funzione Versioni multiple abilitata il file marker `_$folder$` non venisse creato in modo continuo, il che potrebbe portare a un miglioramento delle prestazioni per i bucket con funzione Versioni multiple abilitata.
- Modificato il comportamento in EMRFS in modo tale che non utilizzi i file di istruzione tranne nei casi in cui la crittografia lato client sia abilitata. Se vuoi cancellare i file di istruzione

mentre utilizzi la crittografia lato client, puoi configurare la proprietà `emrfs-site.xml` property, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, su `true`.

- Modificata l'aggregazione dei log YARN per conservare i log nella destinazione dell'aggregazione per due giorni. La destinazione di default è l'archiviazione HDFS del cluster. Se desideri modificare la durata, modifica il valore di `yarn.log-aggregation.retain-seconds` utilizzando la classificazione della configurazione `yarn-site` durante la creazione del cluster. Come sempre, puoi salvare i log dell'applicazione in Amazon S3 utilizzando il parametro `log-uri` quando crei il cluster.

Patch applicate

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.2.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.4.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.7.2	Raccoglitore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccoglitore di metadati Ganglia.
hadoop-client	2.7.1-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.1-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.

Componente	Versione	Descrizione
hadoop-hdfs-library	2.7.1-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.1-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.7.1-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.1-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.1-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.7.1-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.1-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hcatalog-client	1.0.0-amzn-3	Client a riga di comando "hcat" per la manipolazione del server hcatalog.
hcatalog-server	1.0.0-amzn-3	Servizio che fornisce HCatalog, un livello di gestione di tabelle e storage per le applicazioni distribuite.

Componente	Versione	Descrizione
hcatalog-webhcat-server	1.0.0-amzn-3	Endpoint HTTP che fornisce un'interfaccia REST a HCatalog.
hive-client	1.0.0-amzn-3	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-3	Servizio per l'accesso al metastore Hive, un repositor y semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-3	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-5	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.1	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
presto-coordinator	0.136	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.136	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.0	Client a riga di comando Spark.
spark-history-server	1.6.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.0	Librerie Apache Spark necessarie per gli slave YARN.
sqoop-client	1.4.6	Client a riga di comando Apache Sqoop.
webserver	2.4	Server Apache HTTP.
zeppelin-server	0.5.6-incubating	Notebook basato sul Web che consente l'analisi di dati interattiva.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.4.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hcatalog-env	Modifica i valori nell'ambiente di HCatalog.
hcatalog-server-jndi	Modifica i valori nel file jndi.properties di HCatalog.
hcatalog-server-proto-hive-site	Modifica i valori nel file proto-hive-site.xml di HCatalog.
hcatalog-webhcat-env	Modifica i valori nell'ambiente di WebHCat HCatalog.
hcatalog-webhcat-log4j	Modifica i valori nel file log4j.properties di WebHCat HCatalog.
hcatalog-webhcat-site	Modifica i valori nel file webhcat-site.xml file di HCatalog WebHCat.
hive-env	Modifica i valori nell'ambiente Hive.

Classificazioni	Descrizione
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.

Classificazioni	Descrizione
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
sqoop-env	Modifica i valori nell'ambiente di Sqoop.
sqoop-oraoop-site	Modifica i valori nel file oraoop-site.xml di Sqoop OraOop.
sqoop-site	Modifica i valori nel file sqoop-site.xml di Sqoop.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Rilascio 4.3.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
SDK AWS per Java	1,10,27	1,10,27	Non disponibile	Non disponibile
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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
Sqoop	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,5	0,5,5	0.6.0-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.3.0.

Data del rilascio: 19 gennaio 2016

Funzionalità

- Aggiornato a Hadoop 2.7.1
- Aggiornato a Spark 1.6.0
- Aggiornato Ganglia a 3.7.2
- Aggiornato Presto a 0.130
- Amazon EMR ha effettuato alcune modifiche a `spark.dynamicAllocation.enabled` quando è impostato su `true` (vero); è `false` (falso) per impostazione predefinita. Quando è impostato su `true`, interessa i valori predefiniti definiti dall'impostazione `maximizeResourceAllocation`:
 - Se `spark.dynamicAllocation.enabled` è impostato su `true`, `spark.executor.instances` non è impostato da `maximizeResourceAllocation`.

- L'impostazione `spark.driver.memory` ora è configurata in base ai tipi di istanza nel cluster in modo simile a come è impostato `spark.executors.memory`. Tuttavia, dato che l'applicazione del driver di Spark potrebbe essere eseguita o sull'istanza master o su una delle istanze core (ad esempio, nelle modalità client YARN e cluster, rispettivamente), l'impostazione `spark.driver.memory` è definita in base al tipo di istanza del tipo di istanza più piccola tra questi due gruppi di istanze.
- L'impostazione `spark.default.parallelism` è ora definita al doppio dei core CPU disponibili per i container YARN. Nei rilasci precedenti, era la metà del valore.
- I calcoli per il sovraccarico della memoria allocata ai processi YARN di Spark sono stati modificati per essere più precisi, generando un piccolo aumento della quantità di memoria disponibile per Spark (ossia, `spark.executor.memory`).

Problemi noti risolti dai rilasci precedenti

- L'aggregazione dei log YARN ora è abilitata per impostazione predefinita.
- Risolto un problema in cui i log non venivano inviati al bucket di log Amazon S3 del cluster quando l'aggregazione dei log YARN era abilitata.
- Le dimensioni del container YARN ora sono di un minimo di 32 su tutti i tipi di nodo.
- È stato risolto un problema con Ganglia che provocava un I/O su disco eccessivo sul nodo primario nei cluster di grandi dimensioni.
- Risolto un problema che evitava che i log delle applicazioni venissero trasmessi ad Amazon S3 quando un cluster veniva disattivato.
- Risolto un problema in EMRFS CLI che faceva sì che alcuni comandi restituissero errori.
- Risolto un problema con Zeppelin che evitava che le dipendenze venissero caricate nello SparkContext sottostante.
- Risolto un problema causato dall'emissione di un tentativo di ridimensionamento per aggiungere le istanze.
- Risolto un problema in Hive dove `CREATE TABLE AS SELECT` effettua troppe chiamate dell'elenco su Amazon S3.
- Risolto un problema in cui i cluster di grandi dimensioni non effettuano correttamente il provisioning quando sono installati Hue, Oozie e Ganglia.
- Risolto un problema in `s3-dist-cp` dove avrebbe restituito un codice di uscita diverso zero se avesse restituito un errore.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.0.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.1.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.3.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-metadata-collector	3.7.2	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.
ganglia-web	3.7.1	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.7.1-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.7.1-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.7.1-amzn-0	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.7.1-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httpfs-server	2.7.1-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.7.1-amzn-0	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.7.1-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.

Componente	Versione	Descrizione
hadoop-yarn-nodemanager	2.7.1-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.7.1-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hive-client	1.0.0-amzn-2	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-2	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-2	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-5	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.0	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.

Componente	Versione	Descrizione
presto-coordinator	0.130	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.130	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.6.0	Client a riga di comando Spark.
spark-history-server	1.6.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.6.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.6.0	Librerie Apache Spark necessarie per gli slave YARN.
webserver	2.4	Server Apache HTTP.
zeppelin-server	0.5.5-incubating-amzn-1	Notebook basato sul Web che consente l'analisi di dati interattiva.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.3.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file <code>hive-exec-log4j.properties</code> di Hive.
hive-log4j	Modifica i valori nel file <code>hive-log4j.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive.
hue-ini	Modifica i valori nel file <code>ini</code> di Hue.
httpfs-env	Modifica i valori nell'ambiente HTTPFS.

Classificazioni	Descrizione
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.

Classificazioni	Descrizione
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Rilascio 4.2.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)

- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
SDK AWS per Java	1,10,27	1,10,27	Non disponibile	Non disponibile
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Note di rilascio

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.2.0.

Data del rilascio: 18 novembre 2015

Funzionalità

- Aggiunto supporto per Ganglia
- Aggiornato a Spark 1.5.2
- Aggiornato a Presto 0.125
- Aggiornato Oozie a 4.2.0
- Aggiornato Zeppelin a 0.5.5
- Aggiornato AWS SDK for Java a 1.10.27

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con EMRFS CLI dove non utilizzava il nome della tabella dei metadata di default.
- Risolto un problema che avveniva durante l'uso delle tabelle supportate da ORC in Amazon S3.
- Risolto un problema che avveniva con una mancanza di corrispondenza della versione di Python nella configurazione di Spark.
- Risolto un problema quando lo stato di un nodo YARN non riusciva a effettuare il report a causa di problemi di DNS per i cluster in un VPC.
- Risolto un problema che si verificava quando YARN disattivava i nodi, generando applicazioni bloccate o impedendo di programmare nuove applicazioni.
- Risolto un problema che si verificava quando i cluster terminavano con lo stato TIMED_OUT_STARTING.
- Risolto un problema che si verificava quando si include la dipendenza di EMRFS Scala in altre compilazioni. La dipendenza da Scala è stata rimossa.

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I

pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* inizia da 0. Ad esempio, se un componente della community open source denominato myapp-component con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come 2.2-amzn-2.

Componente	Versione	Descrizione
emr-ddb	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.0.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.2.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
ganglia-monitor	3.6.0	Agente Ganglia integrato per le applicazioni dell'ecosistema Hadoop con l'agente di monitoraggio Ganglia.
ganglia-metadata-collector	3.6.0	Raccogliatore di metadati Ganglia per l'aggregazione di parametri degli agenti di monitoraggio Ganglia.

Componente	Versione	Descrizione
ganglia-web	3.5.10	Applicazione Web per la visualizzazione di parametri raccolti dal raccogliatore di metadati Ganglia.
hadoop-client	2.6.0-amzn-2	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.6.0-amzn-2	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.6.0-amzn-2	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.6.0-amzn-2	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.6.0-amzn-2	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.6.0-amzn-2	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.6.0-amzn-2	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.6.0-amzn-2	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.6.0-amzn-2	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hive-client	1.0.0-amzn-1	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-5	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.0	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.2.0	Client a riga di comando Oozie.
oozie-server	4.2.0	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
presto-coordinator	0.125	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.

Componente	Versione	Descrizione
presto-worker	0.125	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.5.2	Client a riga di comando Spark.
spark-history-server	1.5.2	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.5.2	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.5.2	Librerie Apache Spark necessarie per gli slave YARN.
webserver	2.4	Server Apache HTTP.
zeppelin-server	0.5.5-incubating-amzn-0	Notebook basato sul Web che consente l'analisi di dati interattiva.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.2.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file capacity-scheduler.xml di Hadoop.
core-site	Modifica i valori nel file core-site.xml di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.

Classificazioni	Descrizione
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
presto-connector-hive	Modifica i valori nel file hive.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.

Classificazioni	Descrizione
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
spark-metrics	Modifica i valori nel file metrics.properties di Spark.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Rilascio 4.1.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)
- [Classificazioni di configurazione](#)

Versioni delle applicazioni

In questo rilascio sono supportate le seguenti applicazioni: [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
SDK AWS per Java	1,10,27	1,10,27	Non disponibile	Non disponibile
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Note di rilascio

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
<code>emr-ddb</code>	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.
<code>emr-goodies</code>	2.0.0	Librerie utili per l'ecosistema Hadoop.
<code>emr-kinesis</code>	3.1.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
<code>emr-s3-dist-cp</code>	2.0.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
<code>emrfs</code>	2.1.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
hadoop-client	2.6.0-amzn-1	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.6.0-amzn-1	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-library	2.6.0-amzn-1	Libreria e client di riga di comando HDFS
hadoop-hdfs-namenode	2.6.0-amzn-1	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-https-server	2.6.0-amzn-1	Endpoint HTTP per le operazioni HDFS.
hadoop-kms-server	2.6.0-amzn-1	Server di gestione delle chiavi di crittografia basato sull'API KeyProvider di Hadoop.
hadoop-mapred	2.6.0-amzn-1	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.6.0-amzn-1	Servizio YARN per la gestione di container su un singolo nodo.
hadoop-yarn-resourcemanager	2.6.0-amzn-1	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hive-client	1.0.0-amzn-1	Client a riga di comando Hive.

Componente	Versione	Descrizione
hive-metastore-server	1.0.0-amzn-1	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-1	Servizio per l'accettazione di query Hive come richieste Web.
hue-server	3.7.1-amzn-4	Applicazione Web per l'analisi di dati mediante le applicazioni dell'ecosistema Hadoop
mahout-client	0.11.0	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
oozie-client	4.0.1	Client a riga di comando Oozie.
oozie-server	4.0.1	Servizio per l'accettazione delle richieste di flusso di lavoro Oozie.
presto-coordinator	0.119	Servizio per l'accettazione di query e la gestione dell'esecuzione di query di componenti presto-worker.
presto-worker	0.119	Servizio per l'esecuzione di parti di una query.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.5.0	Client a riga di comando Spark.

Componente	Versione	Descrizione
spark-history-server	1.5.0	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.5.0	Motore di esecuzione in memoria per YARN.
spark-yarn-slave	1.5.0	Librerie Apache Spark necessarie per gli slave YARN.
zeppelin-server	0.6.0-incubating-SNAPSHOT	Notebook basato sul Web che consente l'analisi di dati interattiva.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.1.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.

Classificazioni	Descrizione
hadoop-log4j	Modifica i valori nel file log4j.properties di Hadoop.
hdfs-encryption-zones	Configura le zone di crittografia HDFS.
hdfs-site	Modifica i valori nel file hdfs-site.xml di HDFS.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file hive-exec-log4j.properties di Hive.
hive-log4j	Modifica i valori nel file hive-log4j.properties di Hive.
hive-site	Modifica i valori nel file hive-site.xml di Hive
hue-ini	Modifica i valori nel file ini di Hue
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
hadoop-kms-acls	Modifica i valori nel file kms-acls.xml di Hadoop.
hadoop-kms-env	Modifica i valori nell'ambiente Hadoop KMS.
hadoop-kms-log4j	Modifica i valori nel file kms-log4j.properties di Hadoop.
hadoop-kms-site	Modifica i valori nel file kms-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.

Classificazioni	Descrizione
oozie-env	Modifica i valori nell'ambiente di Oozie.
oozie-log4j	Modifica i valori nel file oozie-log4j.properties di Oozie.
oozie-site	Modifica i valori nel file oozie-site.xml di Oozie.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
presto-log	Modifica i valori nel file log.properties di Presto.
presto-config	Modifica i valori nel file config.properties di Presto.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.
zeppelin-env	Modifica i valori nell'ambiente Zeppelin.

Rilascio 4.0.0 di Amazon EMR

- [Versioni delle applicazioni](#)
- [Note di rilascio](#)
- [Versioni dei componenti](#)

- [Classificazioni di configurazione](#)

Versioni delle applicazioni

Questa versione supporta le applicazioni seguenti: [Hadoop](#), [Hive](#), [Mahout](#), [Pig](#) e [Spark](#).

La seguente tabella elenca le versioni dell'applicazione disponibili in questo rilascio di Amazon EMR e quelle nei precedenti tre rilasci di Amazon EMR (quando applicabile).

Per la cronologia completa delle versioni dell'applicazione di ogni rilascio di Amazon EMR, fai riferimento ai seguenti argomenti:

- [Versioni dell'applicazione nei rilasci di Amazon EMR 6.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 5.x](#)
- [Versioni dell'applicazione nei rilasci di Amazon EMR 4.x](#)

Informazioni sulla versione dell'applicazione

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
SDK AWS per Java	1,10,27	1,10,27	Non disponibile	Non disponibile
Python	Non disponibile	Non disponibile	Non disponibile	Non disponibile
Scala	Non disponibile	Non disponibile	Non disponibile	Non disponibile
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

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Note di rilascio

Versioni dei componenti

I componenti che Amazon EMR installa con questo rilascio sono elencati di seguito. Alcuni sono installati come parte di pacchetti di applicazione dei big data. Altri sono specifici per Amazon EMR e installati per processi e caratteristiche del sistema. Questi solitamente iniziano con `emr` o `aws`. I pacchetti di applicazione sui Big data del rilascio di Amazon EMR più recente sono di solito le versioni più recenti reperibili nella community. Mettiamo a disposizione i rilasci della community in Amazon EMR il più rapidamente possibile.

Alcuni componenti in Amazon EMR differiscono dalle versioni della community. Tali componenti hanno un'etichetta che indica la versione nel modulo `CommunityVersion-amzn-EmrVersion`. `EmrVersion` inizia da 0. Ad esempio, se un componente della community open source denominato `myapp-component` con versione 2.2 è stato modificato tre volte per essere incluso in rilasci diversi di Amazon EMR, tale versione di rilascio si presenta come `2.2-amzn-2`.

Componente	Versione	Descrizione
emr-ddb	3.0.0	Connettore di Amazon DynamoDB per le applicazioni dell'ecosistema Hadoop.

Componente	Versione	Descrizione
emr-goodies	2.0.0	Librerie utili per l'ecosistema Hadoop.
emr-kinesis	3.0.0	Connettore di Amazon Kinesis per le applicazioni dell'ecosistema Hadoop.
emr-s3-dist-cp	2.0.0	Applicazione di copia distribuita ottimizzata per Amazon S3.
emrfs	2.0.0	Connettore di Amazon S3 per le applicazioni dell'ecosistema Hadoop.
hadoop-client	2.6.0-amzn-0	Client di riga di comando Hadoop, ad esempio "hdfs", "hadoop" o "yarn".
hadoop-hdfs-datanode	2.6.0-amzn-0	Servizio a livello di nodo HDFS per lo storage di blocchi.
hadoop-hdfs-namenode	2.6.0-amzn-0	Servizio HDFS per tenere traccia dei nomi di file e delle posizioni dei blocchi.
hadoop-httfs-server	2.6.0-amzn-0	Endpoint HTTP per le operazioni HDFS.
hadoop-mapred	2.6.0-amzn-0	Librerie del motore di esecuzione MapReduce per l'esecuzione di un'applicazione MapReduce.
hadoop-yarn-nodemanager	2.6.0-amzn-0	Servizio YARN per la gestione di container su un singolo nodo.

Componente	Versione	Descrizione
hadoop-yarn-resourcemanager	2.6.0-amzn-0	Servizio YARN per l'allocazione e la gestione delle risorse di cluster e delle applicazioni distribuite.
hive-client	1.0.0-amzn-0	Client a riga di comando Hive.
hive-metastore-server	1.0.0-amzn-0	Servizio per l'accesso al metastore Hive, un repository semantico per lo storage di metadati per SQL sulle operazioni Hadoop.
hive-server	1.0.0-amzn-0	Servizio per l'accettazione di query Hive come richieste Web.
mahout-client	0.10.0	Librerie per Machine Learning.
mysql-server	5.5	Server di database MySQL.
pig-client	0.14.0-amzn-0	Client a riga di comando Pig.
spark-client	1.4.1	Client a riga di comando Spark.
spark-history-server	1.4.1	Interfaccia utente Web per la visualizzazione di eventi registrati per il ciclo di vita di un'applicazione Spark completata.
spark-on-yarn	1.4.1	Motore di esecuzione in memoria per YARN.

Componente	Versione	Descrizione
spark-yarn-slave	1.4.1	Librerie Apache Spark necessarie per gli slave YARN.

Classificazioni di configurazione

Le classificazioni di configurazione consentono di personalizzare le applicazioni. Esse corrispondono spesso a un file XML di configurazione per l'applicazione, ad esempio `hive-site.xml`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Classificazioni emr-4.0.0

Classificazioni	Descrizione
capacity-scheduler	Modifica i valori nel file <code>capacity-scheduler.xml</code> di Hadoop.
core-site	Modifica i valori nel file <code>core-site.xml</code> di Hadoop.
emrfs-site	Modifica le impostazioni EMRFS.
hadoop-env	Modifica i valori nell'ambiente Hadoop per tutti i componenti Hadoop.
hadoop-log4j	Modifica i valori nel file <code>log4j.properties</code> di Hadoop.
hdfs-site	Modifica i valori nel file <code>hdfs-site.xml</code> di HDFS.
hive-env	Modifica i valori nell'ambiente Hive.
hive-exec-log4j	Modifica i valori nel file <code>hive-exec-log4j.properties</code> di Hive.
hive-log4j	Modifica i valori nel file <code>hive-log4j.properties</code> di Hive.
hive-site	Modifica i valori nel file <code>hive-site.xml</code> di Hive

Classificazioni	Descrizione
httpfs-env	Modifica i valori nell'ambiente HTTPFS.
httpfs-site	Modifica i valori nel file httpfs-site.xml di Hadoop.
mapred-env	Modifica i valori nell'ambiente MapReduce dell'applicazione MapReduce.
mapred-site	Modifica i valori nel file mapred-site.xml dell'applicazione MapReduce.
pig-properties	Modifica i valori nel file pig.properties di Pig.
pig-log4j	Modifica i valori nel file log4j.properties di Pig.
spark	Impostazioni Amazon EMR per Apache Spark.
spark-defaults	Modifica i valori nel file spark-defaults.conf di Spark.
spark-env	Modifica i valori nell'ambiente Spark.
spark-log4j	Modifica i valori nel file log4j.properties di Spark.
yarn-env	Modifica i valori nell'ambiente YARN.
yarn-site	Modifica i valori nel file yarn-site.xml di YARN.

Versioni AMI 2.x e 3.x di Amazon EMR

Note

AWS sta aggiornando la configurazione TLS per gli endpoint API di AWS a una versione minima di TLS 1.2. Le versioni 3.10 e precedenti di Amazon EMR supportano solo connessioni TLS 1.0/1.1. Dopo il 4 dicembre 2023, non potrai creare cluster con Amazon EMR 3.10 e versioni precedenti.

Se utilizzi Amazon EMR 3.10 o versioni precedenti, consigliamo di testare ed eseguire immediatamente la migrazione dei tuoi carichi di lavoro all'ultima versione di Amazon EMR. Per ulteriori informazioni, consulta il [Blog sulla sicurezza AWS](#).

Le versioni 2.x e 3.x di Amazon EMR, denominate AMI versions (Versioni AMI), sono disponibili per le soluzioni preesistenti che le necessitano per motivi di compatibilità. Non è consigliabile creare nuovi cluster o nuove soluzioni con tali versioni. Non dispongono delle caratteristiche delle versioni più recenti e includono pacchetti di applicazioni obsoleti.

Consigliamo di creare soluzioni utilizzando la versione più recente di Amazon EMR.

Esistono differenze significative tra le versioni 2.x e 3.x e le versioni recenti di Amazon EMR. Tali differenze interessano vari aspetti: la creazione e la configurazione di un cluster, porte e struttura delle directory delle applicazioni sul cluster, ecc.

In questa sezione sono riportate le differenze principali per Amazon EMR, nonché specifiche differenze relative alla configurazione e alla gestione delle applicazioni. Non è quindi esaustiva. Se crei e utilizzi cluster nelle versioni 2.x o 3.x, potresti rilevare delle differenze non descritte in questa sezione.

Argomenti

- [Creazione di un cluster con versioni AMI di Amazon EMR precedenti](#)
- [Installazione di applicazioni con versioni AMI di Amazon EMR precedenti](#)
- [Personalizzazione della configurazione di cluster e applicazioni con versioni AMI di Amazon EMR precedenti](#)
- [Specifiche dell'applicazione Hive per versioni AMI di Amazon EMR precedenti](#)
- [Specifiche dell'applicazione HBase per versioni AMI di Amazon EMR precedenti](#)
- [Specifiche dell'applicazione Pig per versioni AMI di Amazon EMR precedenti](#)
- [Specifiche dell'applicazione Spark con versioni AMI di Amazon EMR precedenti](#)
- [Differenze tra l'utility S3DistCp e le versioni AMI di Amazon EMR precedenti](#)

Creazione di un cluster con versioni AMI di Amazon EMR precedenti

Le versioni 2.x e 3.x di Amazon EMR sono identificate dal numero di versione AMI. La versione 4.0.0 di Amazon EMR e quelle successive sono identificate dal numero di versione mediante un'etichetta di

rilascio, ad esempio `emr-5.11.0`. Questa modifica è più evidente quando crei un cluster mediante l'AWS CLI o a livello di codice.

Quando utilizzi AWS CLI per creare un cluster mediante una versione AMI, utilizza l'opzione `--ami-version`, ad esempio, `--ami-version 3.11.0`. Molte opzioni, caratteristiche e applicazioni introdotte in Amazon EMR 4.0.0 e versioni successive non sono disponibili quando specifichi una `--ami-version`. Per ulteriori informazioni, consulta [create-cluster](#) nella Guida di riferimento ai comandi della AWS CLI.

L'esempio seguente di comando dell'AWS CLI avvia un cluster utilizzando una versione AMI.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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"]
```

A livello di codice, tutte le versioni di Amazon EMR utilizzano l'operazione `RunJobFlowRequest` nell'API di EMR per creare cluster. L'esempio seguente di codice Java crea un cluster mediante la versione AMI 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");
```

La chiamata `RunJobFlowRequest` seguente utilizza invece un'etichetta di release:

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("ReleaseLabel Cluster")
    .withReleaseLabel("emr-5.36.1")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge");
```

Configurazione della dimensione di un cluster

Durante l'esecuzione del cluster, Hadoop determina il numero di attività di mappatore e riduttore necessarie per elaborare i dati. I cluster più grandi devono disporre di più attività per un migliore utilizzo delle risorse e un tempo di elaborazione più basso. Di solito, la dimensione di un cluster EMR non cambia durante l'intero ciclo di vita del cluster; il numero di attività è impostato alla creazione del cluster. Quando ridimensioni un cluster in esecuzione, puoi variare l'elaborazione durante l'esecuzione. Pertanto, anziché utilizzare un numero fisso di attività, puoi variare il numero di attività durante il ciclo di vita del cluster. Sono disponibili due opzioni di configurazione per l'impostazione del numero ideale di attività:

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

Puoi impostare entrambe le opzioni nel file `mapred-conf.xml`. Quando invii un processo al cluster, il client di processi verifica il numero totale corrente di slot di mappatura e riduzione disponibili nell'intero cluster. Il client di processi utilizza quindi le seguenti equazioni per impostare il numero di attività:

- `mapred.map.tasks = mapred.map.tasksperslot * slot di mappatura nel cluster`
- `mapred.reduce.tasks = mapred.reduce.tasksperslot * slot di riduzione nel cluster`

Il client di processi legge il parametro `tasksperslot` solo se il numero di attività non è configurato. Puoi sovrascrivere il numero di attività in qualsiasi momento: per tutti i cluster tramite un'operazione di bootstrap oppure individualmente per processo aggiungendo una fase per modificare la configurazione.

Amazon EMR è in grado di gestire gli errori nei nodi attività e continua l'esecuzione del cluster anche se un nodo attività non è più disponibile. Amazon EMR effettua automaticamente il provisioning di ulteriori nodi attività per sostituire quelli con errori.

Puoi avere un numero differente di nodi di task per ogni fase di cluster. Puoi inoltre aggiungere una fase a un cluster in esecuzione per modificare il numero di nodi di task. Poiché, per impostazione predefinita, l'esecuzione sequenziale delle fasi è garantita, puoi specificare il numero di nodi di task in esecuzione per qualsiasi fase.

Installazione di applicazioni con versioni AMI di Amazon EMR precedenti

Quando utilizzi una versione AMI, le applicazioni sono installate in vari modi, ad esempio utilizzando il parametro `NewSupportedProducts` per l'operazione [RunJobFlow](#), mediante le operazioni di bootstrap e utilizzando l'operazione [Step](#).

Personalizzazione della configurazione di cluster e applicazioni con versioni AMI di Amazon EMR precedenti

Amazon EMR versione 4.0.0 ha introdotto un metodo semplificato di configurazione delle applicazioni mediante le classificazioni di configurazione. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#). Quando utilizzi una versione AMI, configuri le applicazioni utilizzando operazioni di bootstrap insieme agli argomenti che passi. Ad esempio, le operazioni di bootstrap `configure-hadoop` e `configure-daemons` impostano le proprietà di ambiente di Hadoop e YARN come `--namenode-heap-size`. Nelle versioni più recenti, queste proprietà sono configurate mediante le classificazioni di configurazione `hadoop-env` e `yarn-env`. Per le operazioni di bootstrap che eseguono configurazioni comuni, consulta il [repository emr-bootstrap-actions su Github](#).

Le tabelle seguenti eseguono la mappatura delle operazioni di bootstrap alle classificazioni di configurazione in versioni più recenti di Amazon EMR.

Hadoop

Nome di file dell'applicazione interessata	Operazione di bootstrap della versione AMI	Classificazione di configurazione
<code>core-site.xml</code>	<code>configure-hadoop -c</code>	<code>core-site</code>
<code>log4j.properties</code>	<code>configure-hadoop -l</code>	<code>hadoop-log4j</code>

Nome di file dell'applicazione interessata	Operazione di bootstrap della versione AMI	Classificazione di configurazione
hdfs-site.xml	configure-hadoop -s	hdfs-site
N/A	n/a	hdfs-encryption-zones
mapred-site.xml	configure-hadoop -m	mapred-site
yarn-site.xml	configure-hadoop -y	yarn-site
httpfs-site.xml	configure-hadoop -t	httpfs-site
capacity-scheduler.xml	configure-hadoop -z	capacity-scheduler
yarn-env.sh	configure-daemons --resourcemanager-opts	yarn-env

Hive

Nome di file dell'applicazione interessata	Operazione di bootstrap della versione AMI	Classificazione di configurazione
hive-env.sh	N/A	hive-env
hive-site.xml	hive-script --install -hive-site \${MY_HIVE_SITE_FILE}	hive-site
hive-exec-log4j.properties	N/A	hive-exec-log4j
hive-log4j.properties	n/a	hive-log4j

EMRFS

Nome di file dell'applicazione interessata	Operazione di bootstrap della versione AMI	Classificazione di configurazione
emrfs-site.xml	configure-hadoop -e	emrfs-site
N/A	s3get -s s3://customer-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (con la nuova impostazione fs.s3.cse.encryptionMaterialsProvider.uri)

Per un elenco di tutte le classificazioni, consulta [Configurazione delle applicazioni](#).

Variabili di ambiente applicazione

Quando si utilizza una versione AMI, uno script `hadoop-user-env.sh` viene utilizzato insieme all'operazione di bootstrap `configure-daemons` per configurare l'ambiente Hadoop. Lo script include le seguenti operazioni:

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

In Amazon EMR versione 4.x, si procede nello stesso modo utilizzando la classificazione di configurazione `hadoop-env`, come mostrato nell'esempio seguente:

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
  },
  "Configurations": [
    {
      "Classification": "export",
      "Properties": {
        "HADOOP_USER_CLASSPATH_FIRST": "true",
        "HADOOP_CLASSPATH": "/path/to/my.jar"
      }
    }
  ]
}
```

```

    }
  ]
}
]

```

Altro esempio: l'utilizzo di `configure-daemons` e il passaggio di `--namenode-heap-size=2048` e `--namenode-opts=-XX:GCTimeRatio=19` equivale alle classificazioni di configurazione esposte di seguito.

```

[
  {
    "Classification":"hadoop-env",
    "Properties":{

    },
    "Configurations":[
      {
        "Classification":"export",
        "Properties":{
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        }
      }
    ]
  }
]

```

Altre variabili di ambiente applicazione non sono più definite in `/home/hadoop/.bashrc`, ma sono principalmente impostate in file `/etc/default` per componente o applicazione, ad esempio `/etc/default/hadoop`. Anche gli script wrapper in `/usr/bin/` installati dai file RPM delle applicazioni possono impostare ulteriori variabili di ambiente prima dell'utilizzo dello script bin effettivo.

Porte di servizio

Quando si utilizza una versione AMI, alcuni servizi utilizzano porte personalizzate.

Modifiche nelle impostazioni delle porte

Impostazione	AMI versione 3.x	Valore di default open source
<code>fs.default.name</code>	<code>hdfs://emrDeterminedIP:9000</code>	default (<code>hdfs://<i>emrDeterminedIP</i> :8020</code>)

Impostazione	AMI versione 3.x	Valore di default open source
dfs.datanode.address	0.0.0.0:9200	valore di default (0.0.0.0:50010)
dfs.datanode.http.address	0.0.0.0:9102	valore di default (0.0.0.0:50075)
dfs.datanode.https.address	0.0.0.0:9402	valore di default (0.0.0.0:50475)
dfs.datanode.ipc.address	0.0.0.0:9201	valore di default (0.0.0.0:50020)
dfs.http.address	0.0.0.0:9101	valore di default (0.0.0.0:50070)
dfs.https.address	0.0.0.0:9202	valore di default (0.0.0.0:50470)
dfs.secondary.http.address	0.0.0.0:9104	valore di default (0.0.0.0:50090)
yarn.nodemanager.address	0.0.0.0:9103	valore di default (\${yarn.nodemanager.hostname}:0)
yarn.nodemanager.localizer.address	0.0.0.0:9033	valore di default (\${yarn.nodemanager.hostname}:8040)
yarn.nodemanager.webapp.address	0.0.0.0:9035	valore di default (\${yarn.nodemanager.hostname}:8042)
yarn.resourcemanager.address	<i>emrDeterminedIP</i> :9022	valore di default (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> :9025	valore di default (\${yarn.resourcemanager.hostname}:8033)

Impostazione	AMI versione 3.x	Valore di default open source
yarn.resourcemanager.resour ce-tracker.address	<i>emrDeterminedIP</i> :9023	valore di default (\${yarn.r esourcemanager.hos tname}:8031)
yarn.resourcemanager.schedu ler.address	<i>emrDeterminedIP</i> :9024	valore di default (\${yarn.r esourcemanager.hos tname}:8030)
yarn.resourcemanag er.webapp.address	0.0.0.0:9026	valore di default (\${yarn.r esourcemanager.hos tname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> :9046	valore di default (no-value)
yarn.resourcemanager.hostna me	0.0.0.0 (predefinito)	<i>emrDeterminedIP</i>

Note

emrDeterminedIP è un indirizzo IP generato da Amazon EMR.

Utenti

Quando si utilizza una versione di AMI, l'utente hadoop esegue tutti i processi e possiede tutti i file. In Amazon EMR versione 4.0.0 e versioni successive, gli utenti esistono a livello di componente e di applicazione.

Sequenza di installazione, artifact installati e posizioni dei file di log

Quando si utilizza una versione AMI, gli artefatti delle applicazioni e le relative directory di configurazione sono installati nella directory `/home/hadoop/application`. Ad esempio, se hai installato Hive, la directory sarebbe `/home/hadoop/hive`. In Amazon EMR 4.0.0 e versioni successive, gli artifact delle applicazioni sono installati nella directory `/usr/lib/application`. Quando si utilizza una versione AMI, i file di log si trovano in varie posizioni. La tabella seguente elenca tali posizioni.

Modifiche nelle posizioni dei log in Amazon S3

Daemon o applicazione	Percorso della directory
instance-state	node/ <i>instance-id</i> /instance-state/
hadoop-hdfs-namenode	daemons/ <i>instance-id</i> /hadoop-hadoop-namenode.log
hadoop-hdfs-datanode	daemons/ <i>instance-id</i> /hadoop-hadoop-datanode.log
hadoop-yarn (ResourceManager)	daemons/ <i>instance-id</i> /yarn-hadoop-resource-manager
hadoop-yarn (server proxy)	daemons/ <i>instance-id</i> /yarn-hadoop-proxy-server
mapred-historyserver	daemons/ <i>instance-id</i> /
httpsfs	daemons/ <i>instance-id</i> /httpsfs.log
hive-server	node/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	node/ <i>instance-id</i> /apps/hive.log
CLI Hive	node/ <i>instance-id</i> /apps/hive.log
Log utente e log di container delle applicazioni YARN	task-attempts/
Mahout	N/D
Pig	N/D
spark-historyserver	N/D
File della cronologia dei processi mapreduce	jobs/

Runner di comandi

Quando utilizzi una versione AMI, molti script o programmi, come `/home/hadoop/contrib/streaming/hadoop-streaming.jar` non sono posizionati nell'ambiente del percorso di accesso alla shell, pertanto devi specificare il percorso completo quando utilizzi un file jar come `command-runner.jar` o `script-runner.jar` per eseguire gli script. Il file `command-runner.jar` si trova nell'AMI, non è quindi necessario disporre di un URI completo come avveniva con `script-runner.jar`.

Fattore di replica

Il fattore di replica ti consente di stabilire quando avviare una JVM Hadoop. Puoi avviare una nuova JVM Hadoop per ogni attività, in modo da fornire un migliore isolamento delle attività, oppure condividere JVM tra attività, per ridurre l'overhead del framework. Se stai elaborando molti file di piccole dimensioni, è logico riutilizzare la JVM molte volte per ammortizzare il costo di startup. Tuttavia, se ogni attività richiede molto tempo o elabora una grande quantità di dati, puoi scegliere di non riutilizzare la JVM per assicurarti che tutta la memoria sia disponibile per le attività successive. Quando utilizzi una versione AMI, puoi personalizzare il fattore di replica utilizzando l'operazione di `bootstrap configure-hadoop` per impostare la proprietà `mapred.job.reuse.jvm.num.tasks`.

L'esempio seguente illustra l'impostazione del fattore di riutilizzo della JVM per un riutilizzo all'infinito della stessa.

Note




I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).


```
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"]
```

Specifiche dell'applicazione Hive per versioni AMI di Amazon EMR precedenti

File di log

Con le versioni AMI 2.x e 3.x di Amazon EMR, i log Hive sono salvati in `/mnt/var/log/apps/`. Per supportare versioni Hive simultanee, la versione di Hive che esegui determina il nome di file di log, come illustrato nella seguente tabella.

Versione di Hive	Nome del file di log
0.13.1	<code>hive.log</code> <div><p> Note</p><p>A partire da questa versione, Amazon EMR utilizza un nome di file senza versione, ossia <code>hive.log</code>. Le versioni minori condividono la stessa posizione di log della versione principale.</p></div>
0.11.0	<code>hive_0110.log</code> <div><p> Note</p><p>Le versioni minori di Hive 0.11.0, ad esempio 0.11.0.1, condividono la stessa posizione di file di log di Hive 0.11.0.</p></div>
0.8.1	<code>hive_081.log</code> <div><p> Note</p><p>Le versioni minori di Hive 0.8.1, ad esempio Hive 0.8.1.1, condividono la stessa posizione di file di log di Hive 0.8.1.</p></div>
0.7.1	<code>hive_07_1.log</code>

Versione di Hive	Nome del file di log
	<div style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; background-color: #e6f2ff;"> <p> Note</p> <p>Le versioni minori di Hive 0.7.1, ad esempio Hive 0.7.1.3 e Hive 0.7.1.4, condividono la stessa posizione di file di log di Hive 0.7.1.</p> </div>
0.7	hive_07.log
0,5	hive_05.log
0.4	hive.log

Funzionalità di input con suddivisione

Per implementare la funzionalità di input con suddivisione utilizzando le versioni Hive precedenti alla versione 0.13.1 (versioni AMI di Amazon EMR antecedenti alla versione 3.11.0), utilizza quanto segue:

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

Questa funzionalità è stata deprecata a partire dalla versione di Hive 0.13.1. Per utilizzare la stessa funzionalità di input con suddivisione nella versione AMI 3.11.0 di Amazon EMR, utilizza quanto segue:

```
set hive.hadoop.supports.plittable.combineinputformat=true;
```

Porte dei servizi Thrift

Thrift è un framework RPC che definisce un formato di serializzazione binario compatto utilizzato per conservare le strutture di dati per analisi successive. In genere, Hive configura il server per il funzionamento sulle porte seguenti.

Versione di Hive	Numero della porta
Hive 0.13.1	10000
Hive 0.11.0	10004
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001
Hive 0.5	10000

Per ulteriori informazioni sui servizi Thrift, consulta <http://wiki.apache.org/thrift/>.

Utilizzo di Hive per il recupero di partizioni

Amazon EMR include un'istruzione nel linguaggio di query Hive che consente di recuperare le partizioni di una tabella dai dati di tabella in Amazon S3. L'esempio seguente ne è un'illustrazione.

```
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;
```

Le directory e i dati delle partizioni devono trovarsi nella posizione specificata nella definizione di tabella e devono essere denominati in base alla convenzione Hive, ad esempio, dt=2009-01-01.

Note

Dopo Hive 0.13.1, questa funzionalità è supportata in modo nativo mediante `msck repair table` e di conseguenza il supporto di `recover partitions` non è più fornito. Per ulteriori informazioni, consulta <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>.

Passaggio di una variabile Hive a uno script

Per passare una variabile in una fase Hive utilizzando AWS CLI, digitare il comando seguente, sostituire *myKey* con il nome della coppia di chiavi EC2 e sostituire *mybucket* con il nome di bucket. In questo esempio, SAMPLE è un valore di variabile preceduto dal parametro -d. Questa variabile è definita nello script Hive come: `${SAMPLE}`.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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/]
```

Configurazione della posizione di un metastore esterno

La procedura seguente mostra come sovrascrivere i valori di configurazione predefiniti per la posizione del metastore Hive e avviare un cluster utilizzando la posizione del metastore riconfigurato.

Per creare un metastore situato al di fuori del cluster EMR

1. Crea un database MySQL o Aurora utilizzando Amazon RDS.

Per informazioni su come creare un database Amazon RDS, consulta [Nozioni di base su Amazon RDS](#).

2. Modificare i gruppi di sicurezza per autorizzare le connessioni JDBC tra il database e il gruppo di sicurezza ElasticMapReduce-Master.

Per ulteriori informazioni su come modificare i gruppi di sicurezza per l'accesso, consulta la sezione relativa ai [Gruppi di sicurezza Amazon RDS](#) nella Guida per l'utente di Amazon RDS.

3. Impostare i valori di configurazione JDBC in `hive-site.xml`:

- a. Creare un file di configurazione `hive-site.xml` contenente quanto segue:

```
<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 è l'indirizzo DNS dell'istanza Amazon RDS che esegue il database. *username* e *password* sono le credenziali per il database. Per ulteriori informazioni sulla connessione a istanze database MySQL e Aurora, consulta l'argomento relativo alla [Connessione a un'istanza database che esegue il motore del database di MySQL](#) e alla [Connessione a un cluster di database Aurora](#) nella Guida per l'utente di Amazon RDS.

I driver JDBC sono installati da Amazon EMR.

Note

La proprietà `value` non deve contenere spazi o ritorni a capo. Deve essere visualizzato tutto su una riga.

- b. Salvare il file `hive-site.xml` in un percorso su Amazon S3, ad esempio `s3://mybucket/hive-site.xml`.
4. Crea un cluster, specificando il percorso in Amazon S3 del file `hive-site.xml` personalizzato.

L'esempio di comando seguente mostra un comando dell'AWS CLI che esegue questa operazione.

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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"]
```

Connessione ad Hive mediante JDBC

Per eseguire la connessione a Hive mediante JDBC devi scaricare il driver JDBC e installare un client SQL. L'esempio seguente illustra l'utilizzo di SQL Workbench/J per la connessione a Hive mediante JDBC.

Per scaricare i driver JDBC

1. Scaricare ed estrarre i driver appropriati nelle versioni di Hive a cui si desidera accedere. La versione di Hive differisce a seconda dell'AMI scelta alla creazione di un cluster Amazon EMR.
 - Driver JDBC per Hive 0.13.1: https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC_1.0.4.1004.zip
 - Driver JDBC per Hive 0.11.0: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
 - Driver JDBC per Hive 0.8.1: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
2. Installare SQL Workbench/J. Per ulteriori informazioni, consulta la pagina relativa all'[installazione e all'avvio di SQL Workbench/J](#) nella guida per l'utente di SQL Workbench/J.

3. Creare un tunnel SSH per il nodo master del cluster. La porta per la connessione è differente a seconda della versione di Hive. Gli esempi forniti nelle tabelle esposte di seguito sono relativi ai comandi ssh di Linux e ai comandi PuTTY di Windows.

Comandi SSH Linux

Versione di Hive	Comando
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>

Impostazioni del tunnel PuTTY (Windows)

Versione di Hive	Impostazione tunnel
0.13.1	Source port (Porta sorgente): 10000 Destination (Destinazione): <i>master-public-dns-name</i> :10000
0.11.0	Source port (Porta sorgente): 10004 Destination (Destinazione): <i>master-public-dns-name</i> :10004

Versione di Hive	Impostazione tunnel
0.8.1	Source port (Porta sorgente): 10003 Destination (Destinazione): <i>master-public-dns-name</i> :10003

4. Aggiungere il driver JDBC a SQL Workbench.
 - a. Nella finestra di dialogo Select Connection Profile (Seleziona profilo di connessione), scegliere Manage Drivers (Gestisci driver).
 - b. Scegliere l'icona Create a new entry (Crea nuova voce) (pagina bianca).
 - c. Nel campo Nome digitare **Hive JDBC**.
 - d. Per Library (Libreria), fare clic sull'icona Select the JAR file(s) (Seleziona file JAR).
 - e. Selezionare i file JAR come illustrato nella tabella seguente.

Versione driver Hive	File JAR da aggiungere
0.13.1	<pre>hive_metastore.jar hive_service.jar HiveJDBC3.jar libfb303-0.9.0.jar libthrift-0.9.0.jar log4j-1.2.14.jar ql.jar slf4j-api-1.5.8.jar slf4j-log4j12-1.5.8.jar TCLIServiceClient.jar</pre>
0.11.0	<pre>hadoop-core-1.0.3.jar hive-exec-0.11.0.jar hive-jdbc-0.11.0.jar hive-metastore-0.11.0.jar hive-service-0.11.0.jar libfb303-0.9.0.jar commons-logging-1.0.4.jar</pre>

Versione driver Hive	File JAR da aggiungere
	<code>slf4j-api-1.6.1.jar</code>
0.8.1	<code>hadoop-core-0.20.205.jar</code> <code>hive-exec-0.8.1.jar</code> <code>hive-jdbc-0.8.1.jar</code> <code>hive-metastore-0.8.1.jar</code> <code>hive-service-0.8.1.jar</code> <code>libfb303-0.7.0.jar</code> <code>libthrift-0.7.0.jar</code> <code>log4j-1.2.15.jar</code> <code>slf4j-api-1.6.1.jar</code> <code>slf4j-log4j12-1.6.1.jar</code>
0.7.1	<code>hadoop-0.20-core.jar</code> <code>hive-exec-0.7.1.jar</code> <code>hive-jdbc-0.7.1.jar</code> <code>hive-metastore-0.7.1.jar</code> <code>hive-service-0.7.1.jar</code> <code>libfb303.jar</code> <code>commons-logging-1.0.4.jar</code> <code>slf4j-api-1.6.1.jar</code> <code>slf4j-log4j12-1.6.1.jar</code>
0.7	<code>hadoop-0.20-core.jar</code> <code>hive-exec-0.7.0.jar</code> <code>hive-jdbc-0.7.0.jar</code> <code>hive-metastore-0.7.0.jar</code> <code>hive-service-0.7.0.jar</code> <code>libfb303.jar</code> <code>commons-logging-1.0.4.jar</code> <code>slf4j-api-1.5.6.jar</code> <code>slf4j-log4j12-1.5.6.jar</code>

Versione driver Hive	File JAR da aggiungere
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. Nella finestra di dialogo Please select one driver (Selezionare un driver), selezionare un driver in base alla tabella riportata di seguito e fare clic su OK.

Versione di Hive	Nome di classe del driver
0.13.1	<code>com.amazon.hive.jdbc3.HS2Driver</code>
0.11.0	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.8.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0,5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. Quando ritorni alla finestra di dialogo Select Connection Profile (Seleziona profilo di connessione), verifica che il campo Driver sia impostato su Hive JDBC (JDBC Hive) e fornisci la stringa di connessione JDBC nel campo URL in base alla tabella seguente.

Versione di Hive	Stringa di connessione JDBC
0.13.1	<code>jdbc:hive2://localhost:10000/default</code>
0.11.0	<code>jdbc:hive://localhost:10004/default</code>
0.8.1	<code>jdbc:hive://localhost:10003/default</code>

Se il cluster utilizza AMI versione 3.3.1 o successiva, nella finestra di dialogo Select Connection Profile (Seleziona profilo di connessione), digitare **hadoop** nel campo Username (Nome utente).

Specifiche dell'applicazione HBase per versioni AMI di Amazon EMR precedenti

Versioni di HBase supportate

Versione di HBase	Versione AMI	Parametri di configurazione AWS CLI	Dettagli della versione di HBase
0.94.18	versioni 3.1.0 e successive	<pre>--ami-version 3.1 --ami-version 3.2 --ami-version 3.3 --applications Name=HBase</pre>	<ul style="list-style-type: none"> Miglioramenti e correzioni di bug.
0.94.7	3.0-3.0.4	<pre>--ami-version 3.0</pre>	

Versione di HBase	Versione AMI	Parametri di configurazione AWS CLI	Dettagli della versione di HBase
0.92	versioni 2.2 e successive	<pre>--applications Name=HBase --ami-version 2.2 or later --applications Name=HBase</pre>	

Prerequisiti per i cluster HBase

Un cluster creato utilizzando le versioni AMI 2.x e 3.x di Amazon EMR deve soddisfare i seguenti requisiti per HBase.

- L'AWS CLI (facoltativo) - Per interagire con HBase utilizzando la riga di comando, scaricare e installare la versione più recente dell'AWS CLI. Per ulteriori informazioni, consulta [Installazione dell'AWS Command Line Interface](#) nella Guida per l'utente di AWS Command Line Interface.
- Almeno due istanze (facoltativo): il nodo master del cluster esegue il server master HBase e Zookeeper mentre i nodi attività eseguono i server di regione HBase. Per le migliori prestazioni, i cluster HBase devono essere eseguiti su almeno due istanze EC2, ma è possibile eseguire HBase su un singolo nodo per scopi di valutazione.
- Cluster di lunga durata: HBase viene eseguito soltanto su cluster di lunga durata. Per impostazione predefinita, la CLI e la console di Amazon EMR creano cluster di lunga durata.
- Una coppia di chiavi Amazon EC2 impostata (consigliato): per utilizzare il protocollo di rete SSH (Secure Shell) per la connessione al nodo master e l'esecuzione dei comandi della shell HBase, è necessario utilizzare un coppia di chiavi Amazon EC2 alla creazione del cluster.
- Versioni AMI e Hadoop corrette - I cluster HBase sono attualmente supportati solo su Hadoop 20.205 o versioni successive.
- Ganglia (facoltativo) - Per monitorare i parametri di prestazioni HBase, installare Ganglia alla creazione del cluster.
- Un bucket Amazon S3 per i log (facoltativo): i log di HBase sono disponibili sul nodo master. Per copiare questi log in Amazon S3, specifica un bucket S3 per ricevere i file di log alla creazione del cluster.

Creazione di un cluster con HBase

La tabella seguente elenca le opzioni disponibili quando si utilizza la console per creare un cluster con HBase mediante una versione AMI di Amazon EMR.

Campo	Azione
Restore from backup (Ripristina da backup)	Specifica se precaricare il cluster HBase con i dati archiviati in Amazon S3.
Backup location (Posizione backup)	Specifica l'URI in cui il backup per il ripristino risiede in Amazon S3.
Backup version (Versione backup)	Eventualmente, specificare il nome della versione del backup in Backup Location (Posizione backup) che si intende utilizzare. Se si lascia questo campo vuoto, Amazon EMR utilizza il backup più recente in Backup Location (Percorso backup) per popolare il nuovo cluster HBase.
Schedule Regular Backups (Pianifica backup regolari)	Specificare se pianificare backup incrementali automatici. Il primo backup è un backup completo per creare una baseline per i backup incrementali successivi.
Consistent backup (Backup coerente)	Specificare se i backup devono essere coerenti. Un backup coerente è un backup che sospende le operazioni di scrittura durante la fase di backup iniziale, ovvero la sincronizzazione dei nodi. Tutte le operazioni di scrittura sospese sono inserite in una coda e ripristinate al termine della sincronizzazione.
Backup frequency (Frequenza backup)	Il numero di giorni/ore/minuti tra i backup pianificati.

Campo	Azione
Backup location (Posizione backup)	L'URI di Amazon S3 in cui vengono archiviati i backup. Il percorso di backup di ogni cluster HBase deve essere differente per assicurare la correttezza dei backup differenziali.
Backup start time (Orario di avvio backup)	Specificare quando deve iniziare il primo backup. Specificare <code>now</code> per iniziare il primo backup non appena il cluster viene eseguito oppure immettere una data e un'ora in formato ISO . Ad esempio, impostando <code>2012-06-15T20:00Z</code> , il backup ha inizio il 15 giugno 2012 alle 20.00 UTC.

L'esempio di comando dell'AWS CLI seguente consente di avviare un cluster con HBase e altre applicazioni:

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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
```

Dopo che la connessione tra i cluster Hive e HBase è stata effettuata (come mostrato nella procedura precedente), puoi accedere ai dati archiviati nel cluster HBase creando una tabella esterna in Hive.

L'esempio seguente, quando eseguito dal prompt Hive, crea una tabella esterna che fa riferimento ai dati archiviati in una tabella HBase denominata `inputTable`. Puoi quindi fare riferimento a `inputTable` nelle istruzioni Hive per eseguire una query e modificare i dati memorizzati nel cluster HBase.

Note

L'esempio seguente utilizza `protobuf-java-2.4.0a.jar` nell'AMI 2.3.3, ma devi modificare l'esempio affinché corrisponda alla versione utilizzata. Per determinare qual è la versione in

uso del file JAR dei buffer di protocollo, esegui il comando dal prompt dei comandi Hive: !
`ls /home/hadoop/lib;`

```
add jar lib/emr-metrics-1.0.jar ;
    add jar lib/protobuf-java-2.4.0a.jar ;

set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com ;

create external table inputTable (key string, value string)
    stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
    tblproperties ("hbase.table.name" = "t1");

select count(*) from inputTable ;
```

Personalizzazione della configurazione di HBase

Anche se la maggior parte delle applicazioni funziona correttamente con le impostazioni predefinite, puoi scegliere di modificare le impostazioni di configurazione HBase. A questo proposito, esegui uno dei due script di operazione di bootstrap:

- `configure-hbase-daemons`: configura le proprietà dei daemon master, regionserver e zookeeper. Queste proprietà includono la dimensione heap e le opzioni per il passaggio alla JVM all'avvio del daemon HBase. Puoi impostare queste proprietà come argomenti nell'operazione di bootstrap. Questa operazione di bootstrap modifica il file di configurazione `/home/hadoop/conf/hbase-user-env.sh` nel cluster HBase.
- `configure-hbase`: configura impostazioni specifiche al sito HBase, ad esempio la porta a cui il master HBase deve associarsi e il numero massimo di tentativi che il client della CLI ha a disposizione per eseguire un'operazione. Puoi impostare questi parametri uno a uno, come argomenti nell'operazione di bootstrap, oppure specificare il percorso di un file di configurazione XML in Amazon S3. Questa operazione di bootstrap modifica il file di configurazione `/home/hadoop/conf/hbase-site.xml` nel cluster HBase.

Note

Questi script, come altre operazioni di bootstrap, possono essere eseguiti solo alla creazione del cluster; non puoi utilizzarli per modificare la configurazione di un cluster HBase in esecuzione.

Quando esegui le operazioni di bootstrap `configure-hbase` o `configure-hbase-daemon`, i valori specificati sostituiscono i valori predefiniti. Ai parametri non impostati in modo esplicito sono assegnati i valori di default.

La configurazione di HBase con queste operazioni di bootstrap è analoga all'utilizzo di operazioni di bootstrap in Amazon EMR per configurare le impostazioni Hadoop e le proprietà di daemon di Hadoop. La differenza è che HBase non ha opzioni di memoria per processo. Le opzioni di memoria sono invece impostate utilizzando l'argomento `--daemon-opts`, dove *daemon* è sostituito dal nome del daemon da configurare.

Configurazione di daemon HBase

Amazon EMR fornisce un'operazione di bootstrap, `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, che puoi utilizzare per modificare la configurazione dei daemon HBase, dove *region* è la regione in cui il cluster HBase è avviato.

Per configurare i daemon HBase utilizzando AWS CLI, aggiungi l'operazione di bootstrap `configure-hbase-daemons` all'avvio del cluster per configurare uno o più daemon HBase. Puoi impostare le seguenti proprietà:

Proprietà	Descrizione
<code>hbase-master-opts</code>	Opzioni che controllano il modo in cui la JVM esegue il daemon master. Se impostate, sovrascrivono le variabili <code>HBASE_MASTER_OPTS</code> di default.
<code>regionserver-opts</code>	Opzioni che controllano il modo in cui la JVM esegue il daemon del server di regione. Se impostate, sovrascrivono le variabili <code>HBASE_REGIONSERVER_OPTS</code> di default.

Proprietà	Descrizione
<code>zookeeper-opts</code>	Opzioni che controllano il modo in cui la JVM esegue il daemon zookeeper. Se impostate, sovrascrivono le variabili <code>HBASE_ZOOKEEPER_OPTS</code> di default.

Per ulteriori informazioni su queste opzioni, consulta [hbase-env.sh](#) nella documentazione di HBase.

Un'operazione di bootstrap per configurare i valori per `zookeeper-opts` e `hbase-master-opts` è illustrata nell'esempio seguente.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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"]
```

Configurazione delle impostazioni del sito HBase

Amazon EMR fornisce un'operazione di bootstrap, `s3://elasticmapreduce/bootstrap-actions/configure-hbase`, che puoi utilizzare per modificare la configurazione di HBase. Puoi impostare questi valori di configurazione uno a uno, come argomenti nell'operazione di bootstrap, oppure specificare il percorso di un file di configurazione XML in Amazon S3. L'impostazione di ogni singolo valore di configurazione è utile se è necessario impostare solo alcune impostazioni di configurazione. L'impostazione mediante un file XML è utile se devi apportare numerose modifiche oppure se intendi salvare le impostazioni di configurazione per un riutilizzo futuro.

Note

Puoi prefissare il nome del bucket Amazon S3 con il prefisso di una regione, ad esempio `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`, dove *region* è la regione in cui si avvia il cluster HBase.

Questa operazione di bootstrap modifica il file di configurazione `/home/hadoop/conf/hbase-site.xml` nel cluster HBase. L'operazione di bootstrap può essere eseguita solo all'avvio del cluster HBase.

Per ulteriori informazioni sulle impostazioni del sito HBase che è possibile configurare, consulta la sezione relativa alla [Default configuration \(Configurazione predefinita\)](#) nella documentazione di HBase.

Imposta l'operazione di bootstrap `configure-hbase` quando avvii il cluster HBase e specifica i valori da modificare in `hbase-site.xml`.

Per specificare singole impostazioni del sito HBase mediante l'AWS CLI

- Per modificare l'impostazione `hbase.hregion.max.filesize`, digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi Amazon EC2.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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"]
```

Per specificare le impostazioni del sito HBase con un file XML mediante l'AWS CLI

1. Creare una versione personalizzata di `hbase-site.xml`. Il file personalizzato devono essere un file XML valido. Per ridurre il rischio di errori, esegui l'avvio con la copia predefinita di `hbase-site.xml`, disponibile nel nodo master HBase di Amazon EMR in `/home/hadoop/conf/hbase-site.xml`, quindi modifica una copia di quel file anziché creare un file da zero. È possibile assegnare un nuovo nome al file oppure mantenere il nome `hbase-site.xml`.
2. Carica il file `hbase-site.xml` personalizzato in un bucket Amazon S3. Questo file deve disporre di autorizzazioni che consentano all'account AWS che avvia il cluster di accedervi. Se l'account AWS che avvia il cluster è anche il proprietario del bucket Amazon S3, ha accesso al file.
3. Impostare l'operazione di bootstrap `configure-hbase` quando si avvia il cluster HBase e includere il percorso del file `hbase-site.xml` personalizzato. L'esempio seguente imposta i valori di configurazione del sito HBase su quelli specificati nel file `s3://mybucket/my-hbase-site.xml`. Digita il comando seguente, sostituisci *myKey* con il nome della coppia di chiavi EC2 e sostituisci *mybucket* con il nome del bucket Amazon S3.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
  --applications Name=Hue Name=Hive Name=Pig Name=HBase \
  --use-default-roles --ec2-attributes KeyName=myKey \
  --instance-type c1.xlarge --instance-count 3 --termination-protected \
  --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["--site-config-file","s3://mybucket/config.xml"]
```

Se si specifica più di un'opzione per personalizzare il funzionamento di HBase, è necessario anteporre a ogni coppia chiave-valore un parametro di opzione `-s` come mostrato nell'esempio seguente:

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s","zookeeper.session.timeout=60000"]
```

Con il proxy impostato e la connessione SSH aperta, è possibile visualizzare l'interfaccia utente HBase aprendo una finestra del browser con `http://master-public-dns-name:60010/master-status`, dove *master-public-dns-name* è l'indirizzo DNS pubblico del nodo master nel cluster HBase.

È possibile visualizzare i log di HBase correnti utilizzando SSH per la connessione al nodo master e accedendo alla directory `mnt/var/log/hbase`. Questi file di log non sono disponibili dopo che il cluster viene terminato a meno che non si abiliti la registrazione in Amazon S3 quando il cluster viene avviato.

Backup e ripristino di HBase

Amazon EMR fornisce la possibilità di eseguire, manualmente o mediante una pianificazione automatizzata, il backup dei dati HBase in Amazon S3. Puoi eseguire backup completi e incrementali. Dal momento in cui disponi di una versione di backup dei dati HBase, puoi ripristinare quella versione in un cluster HBase. Puoi eseguire il ripristino in un cluster HBase in esecuzione oppure avviare un nuovo cluster prepopolato con i dati di backup.

Durante il processo di backup, HBase continua a eseguire i comandi di scrittura. Ciò garantisce la disponibilità del cluster durante il backup, ma anche il rischio di inconsistenze tra i dati sottoposti a backup e le operazioni di scrittura eseguite in parallelo. Per comprendere le incoerenze che potrebbero verificarsi, devi considerare che HBase distribuisce le operazioni di scrittura sui nodi nel relativo cluster. Se un'operazione di scrittura viene eseguita dopo il polling di un determinato nodo, quei dati non sono inclusi nell'archivio di backup. È persino possibile che scritture precedenti sul cluster HBase (inviata a un nodo di cui è già stato eseguito il polling) non siano presenti nell'archivio di backup, mentre quelle successive (inviata a un nodo prima del polling) lo siano.

Se un backup coerente è necessario, devi sospendere le operazioni di scrittura su HBase durante la parte iniziale del processo di backup, ovvero la sincronizzazione dei nodi. Puoi eseguire questa operazione specificando il parametro `--consistent` quando richiedi un backup. Con tale parametro, le operazioni di scrittura in questo periodo vengono messe in coda ed eseguite subito dopo il termine della sincronizzazione. Puoi anche pianificare backup ricorrenti e quindi risolvere eventuali incoerenze nel tempo, in quanto i dati non inclusi in un ciclo di backup sono inclusi in quello successivo.

Quando esegui il backup dei dati HBase, devi specificare una directory di backup differente per ogni cluster. Un modo semplice di eseguire questa operazione è utilizzare l'identificatore del cluster come parte del percorso specificato per la directory di backup. Ad esempio, `s3://mybucket/backups/j-3AEXXXXXX16F2`. Ciò garantisce che tutti i backup incrementali futuri faranno riferimento al cluster HBase corretto.

Quando sei pronto per eliminare vecchi file di backup non più necessari, ti consigliamo di eseguire dapprima un backup completo dei dati HBase. Questa operazione assicura la conservazione di tutti i dati e fornisce una baseline per i backup incrementali successivi. Al termine del backup completo, puoi accedere alla directory di backup ed eliminare manualmente i vecchi file di backup.

Per l'operazione di copia il processo di backup di HBase utilizza S3DistCp, che presenta alcune limitazioni in relazione allo spazio di storage dei file temporanei.

Backup e ripristino di HBase mediante la console

La console fornisce la possibilità di avviare un nuovo cluster e popolarlo con i dati di un backup di HBase precedente. Consente inoltre di pianificare backup incrementali periodici dei dati HBase. Altre funzionalità di backup e ripristino, come la possibilità di ripristinare dati in un cluster già in esecuzione, di eseguire backup manuali e pianificare backup completi automatici, sono disponibili tramite la CLI.

Per popolare un nuovo cluster con dati HBase archiviati utilizzando la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Scegli Crea cluster.
3. Nella sezione Storage Configuration (Configurazione storage), per Additional Applications (Ulteriori applicazioni), scegliere HBase e Configure and add (Configura e aggiungi).
4. Nella finestra di dialogo Add Application (Aggiungi applicazione), selezionare Restore From Backup (Ripristina da backup).
5. Per Backup Location (Percorso backup), specificare il percorso del backup da caricare nel nuovo cluster HBase. Dovrebbe trattarsi di un URL Amazon S3 nel formato `s3://myawsbucket/backups/`.
6. Per Backup Version (Versione backup), si ha possibilità di specificare il nome di una versione di backup da caricare impostando un valore. Se non si imposta un valore per Backup Version (Versione backup), Amazon EMR carica il backup più recente nel percorso specificato.

7. Scegliere Add (Aggiungi) e procedere alla creazione del cluster con altre opzioni, se necessario.

Per pianificare backup automatizzati dei dati HBase mediante la console

1. Nella sezione Storage Configuration (Configurazione storage), per Additional Applications (Ulteriori applicazioni), scegliere HBase e Configure and add (Configura e aggiungi).
2. Scegliere Schedule Regular Backups (Pianifica backup regolari).
3. Specificare se i backup devono essere coerenti. Un backup coerente è un backup che sospende le operazioni di scrittura durante la fase di backup iniziale, ovvero la sincronizzazione dei nodi. Tutte le operazioni di scrittura sospese sono inserite in una coda e ripristinate al termine della sincronizzazione.
4. Impostare la frequenza dei backup immettendo un numero per Backup Frequency (Frequenza backup) e scegliendo Days (Giorni), Hours (Ore) o Minutes (Minuti). Il primo backup automatico che viene eseguito è un backup completo; successivamente, Amazon EMR salva backup incrementali in base alla pianificazione specificata.
5. Specifica il percorso in Amazon S3 in cui i backup devono essere archiviati. Il backup di ogni cluster HBase deve essere salvato in un percorso distinto in Amazon S3 per garantire il calcolo corretto dei backup incrementali.
6. Specificare quando deve verificarsi il primo backup impostando un valore per Backup Start Time (Orario di avvio backup). Specificare now per iniziare il primo backup non appena il cluster viene eseguito oppure immettere una data e un'ora in [formato ISO](#). Ad esempio, impostando 2013-09-26T20:00Z, il backup ha inizio il 26 settembre 2013 alle 20.00 UTC.
7. Scegliere Add (Aggiungi).
8. Procedere con la creazione del cluster con altre opzioni, se necessario.

Monitoraggio di HBase con CloudWatch

Amazon EMR trasmette tre parametri a CloudWatch che puoi utilizzare per monitorare i backup di HBase. Questi parametri sono trasmessi a CloudWatch ogni cinque minuti e sono forniti gratuitamente.

Parametro	Descrizione
HBaseBackupFailed	Se l'ultimo backup non è riuscito. È impostato su 0 per impostazione predefinita e diventa 1 se il tentativo di

Parametro	Descrizione
	<p>backup precedente non riesce. Questo parametro è indicato solo per i cluster HBase.</p> <p>Caso d'uso: monitorare i backup HBase</p> <p>Unità: numero</p>
HBaseMostRecentBackupDuration	<p>Il tempo necessario per il completamento del backup precedente. Questo parametro viene impostato indipendentemente dalla riuscita o meno dell'ultimo backup completato. Durante l'esecuzione del backup, questo parametro restituisce il numero di minuti trascorsi dopo l'avvio del backup. Questo parametro è indicato solo per i cluster HBase.</p> <p>Caso d'uso: monitorare i backup HBase</p> <p>Unità: minuti</p>
HBaseTimeSinceLastSuccessfulBackup	<p>Il numero di minuti trascorsi dopo l'ultimo backup HBase riuscito avviato sul tuo cluster. Questo parametro è indicato solo per i cluster HBase.</p> <p>Caso d'uso: monitorare i backup HBase</p> <p>Unità: minuti</p>

Configurazione di Ganglia per HBase

Puoi configurare Ganglia per HBase utilizzando l'operazione di bootstrap `configure-hbase-for-ganglia`. Questa operazione configura HBase per pubblicare parametri in Ganglia.

Devi configurare HBase e Ganglia quando avvii il cluster; i report Ganglia non possono essere aggiunti a un cluster in esecuzione.

Ganglia archivia anche i file di log sul server all'indirizzo `/mnt/var/log/ganglia/rlds`. Se hai configurato il cluster per mantenere i file di log in un bucket Amazon S3, i file di log Ganglia vengono mantenuti anche in questo percorso.

Per avviare un cluster con Ganglia per HBase, utilizza l'operazione di bootstrap `configure-hbase-for-ganglia` come mostrato nell'esempio seguente.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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
```

Dopo che il cluster viene avviato con Ganglia configurato, puoi accedere ai grafici e ai report Ganglia utilizzando l'interfaccia grafica in esecuzione sul nodo master.

Specifiche dell'applicazione Pig per versioni AMI di Amazon EMR precedenti

Versioni di Pig supportate

La versione di Pig che puoi aggiungere al cluster dipende dalla versione AMI di Amazon EMR e dalla versione Hadoop utilizzate. La tabella seguente mostra quali versioni AMI e versioni di Hadoop sono compatibili con le differenti versioni di Pig. Ti consigliamo di utilizzare la versione disponibile più recente di Pig per sfruttare i miglioramenti delle prestazioni e le nuove funzionalità.

Quando utilizzi l'API per installare Pig, viene utilizzata la versione predefinita a meno che non specifichi `--pig-versions` come argomento per la fase che carica Pig sul cluster durante la chiamata a [RunJobFlow](#).

Versione di Pig	Versione AMI	Parametri di configurazione	Dettagli della versione di Pig
0.12.0	versioni 3.1.0 e successive	<code>--ami-version 3.1</code>	Aggiunge il supporto per:

Versione di Pig	Versione AMI	Parametri di configurazione	Dettagli della versione di Pig
Note di rilascio Documentazione		<pre>--ami-version 3.2 --ami-version 3.3</pre>	<ul style="list-style-type: none"> • Streaming di UDF senza implementazioni JVM • Operatori ASSERT e IN • Espressione CASE • AvroStorage come funzione integrata Pig • ParquetLoader e ParquetStorer come funzioni integrate • Tipi BigInteger e BigDecimal
0.11.1.1 Note di rilascio Documentazione	versioni 2.2 e successive	<pre>--pig-versions 0.11.1.1 --ami-version 2.2</pre>	Migliora le prestazioni del comando LOAD con PigStorage se l'input risiede in Amazon S3.
0.11.1 Note di rilascio Documentazione	versioni 2.2 e successive	<pre>--pig-versions 0.11.1 --ami-version 2.2</pre>	Aggiunge supporto per JDK 7, Hadoop 2, funzioni UDF Groovy, ottimizzazione SchemaTuple, nuovi operatori e altro. Per ulteriori informazioni, consulta Pig 0.11.1 Change Log .

Versione di Pig	Versione AMI	Parametri di configurazione	Dettagli della versione di Pig
0.9.2.2 Note di rilascio Documentazione	versioni 2.2 e successive	--pig-versions 0.9.2.2 --ami-version 2.2	Aggiunge il supporto per Hadoop 1.0.3.
0.9.2.1 Note di rilascio Documentazione	versioni 2.2 e successive	--pig-versions 0.9.2.1 --ami-version 2.2	Aggiunge supporto per MapR.
0.9.2 Note di rilascio Documentazione	versioni 2.2 e successive	--pig-versions 0.9.2 --ami-version 2.2	Include vari miglioramenti delle prestazioni e correzioni di bug. Per informazioni complete sulle modifiche per Pig 0.9.2, consulta Pig 0.9.2 Change Log .
0.9.1 Note di rilascio Documentazione	2.0	--pig-versions 0.9.1 --ami-version 2.0	
0.6 Note di rilascio	1.0	--pig-versions 0.6 --ami-version 1.0	

Versione di Pig	Versione AMI	Parametri di configurazione	Dettagli della versione di Pig
0.3 Note di rilascio	1.0	<pre>--pig-versions 0.3 --ami-version 1.0</pre>	

Dettagli della versione di Pig

Amazon EMR supporta alcune versioni di Pig a cui potrebbero essere state applicate ulteriori patch di Amazon EMR. Puoi configurare la versione di Pig da eseguire su cluster Amazon EMR. Per ulteriori informazioni su come effettuare tale operazione, consulta [Apache Pig](#). Le seguenti sezioni descrivono varie versioni di Pig e le patch applicate alle versioni caricate su Amazon EMR.

Patch Pig

Questa sezione descrive le patch personalizzate applicate alle versioni di Pig disponibili con Amazon EMR.

Patch di Pig 0.11.1.1

La versione Amazon EMR di Pig 0.11.1.1 è una versione di manutenzione che migliora le prestazioni del comando LOAD con PigStorage se l'input risiede in Amazon S3.

Patch di Pig 0.11.1

La versione Amazon EMR di Pig 0.11.1 contiene tutti gli aggiornamenti forniti da Apache Software Foundation e le patch Amazon EMR cumulative di Pig versione 0.9.2.2. Tuttavia, Pig 0.11.1 non include alcuna nuova patch specifica per Amazon EMR.

Patch di Pig 0.9.2

Apache Pig 0.9.2 è una versione di manutenzione di Pig. Il team di Amazon EMR ha applicato le seguenti patch alla versione Amazon EMR di Pig 0.9.2.

Patch	Descrizione
PIG-1429	<p>Aggiunta del tipo di dati booleano a Pig come tipo di dati di prima classe. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-1429.</p> <p>Stato: confermata</p> <p>Corretto in Apache Pig versione: 0.10</p>
PIG-1824	<p>Supporto per moduli di importazione in funzioni UDF Jython. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-1824.</p> <p>Stato: confermata</p> <p>Corretto in Apache Pig versione: 0.10</p>
PIG-2010	<p>File JAR registrati in bundle nella cache distribuita. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-2010.</p> <p>Stato: confermata</p> <p>Corretto in Apache Pig versione: 0.11</p>
PIG-2456	<p>Aggiunta di un file <code>~/.pigbootup</code> in cui l'utente può specificare istruzioni Pig di default. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-2456.</p> <p>Stato: confermata</p> <p>Corretto in Apache Pig versione: 0.11</p>
PIG-2623	<p>Supporto per l'utilizzo di percorsi Amazon S3 per la registrazione di funzioni UDF. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-2623.</p> <p>Stato: confermata</p>

Patch	Descrizione
	Corretto in Apache Pig versione: 0.10, 0.11

Patch di Pig 0.9.1

Il team di Amazon EMR ha applicato le seguenti patch alla versione Amazon EMR di Pig 0.9.1.

Patch	Descrizione
Supporto per file JAR e script Pig in dfs	<p>Aggiunta del supporto per l'esecuzione di script e la registrazione di file JAR archiviati in HDFS, Amazon S3 o altri file system distribuiti. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-1505.</p> <p>Stato: confermata</p> <p>Corretto in Apache Pig versione: 0.8.0</p>
Supporto per più file system in Pig	<p>Aggiunta del supporto per script Pig per la lettura di dati da un file system e la scrittura su un altro. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-1564.</p> <p>Stato: non confermata</p> <p>Corretto in Apache Pig versione: n/d</p>
Aggiunta di funzioni UDF datettime e string Piggybank	<p>Aggiunta di funzioni UDF datettime e string per supportare script Pig personalizzati. Per ulteriori informazioni, consulta la pagina all'indirizzo https://issues.apache.org/jira/browse/PIG-1565.</p> <p>Stato: non confermata</p> <p>Corretto in Apache Pig versione: n/d</p>

Cluster Pig in modalità interattiva o batch

Amazon EMR consente di eseguire script Pig in due modi:

- Interactive
- Archiviazione

Quando avvii un cluster di lunga durata mediante la console o AWS CLI, puoi eseguire la connessione utilizzando ssh nel nodo master come utente Hadoop e utilizzare la shell Grunt per sviluppare ed eseguire script Pig in modalità interattiva. L'utilizzo di Pig in modalità interattiva ti consente di rivedere lo script Pig più facilmente rispetto alle modalità batch. Dopo aver verificato lo script Pig in modalità interattiva, puoi caricarlo in Amazon S3 e utilizzare la modalità batch per eseguire lo script in produzione. Puoi inoltre inviare comandi Pig in modalità interattiva su un cluster in esecuzione per analizzare e trasformare i dati come desiderato.

In modalità batch, carichi lo script Pig in Amazon S3 e invii il lavoro al cluster come fase. Le fasi Pig possono essere inviate a un cluster di lunga durata o a un cluster transitorio.

Specifiche dell'applicazione Spark con versioni AMI di Amazon EMR precedenti

Utilizzo di Spark in modalità interattiva o batch

Amazon EMR ti consente di eseguire le applicazioni Spark in due modalità:

- Interactive
- Archiviazione

Quando avvii un cluster di lunga durata mediante la console o l'AWS CLI, puoi eseguire la connessione utilizzando SSH nel nodo master come utente Hadoop e utilizzare la shell Spark per sviluppare ed eseguire le applicazioni Spark in modalità interattiva. L'utilizzo di Spark in modalità interattiva ti consente di realizzare prototipi di applicazioni Spark o di testarle più facilmente rispetto a un ambiente batch. Dopo aver esaminato senza problemi l'applicazione Spark in modalità interattiva, puoi inserire il programma JAR o Python di quell'applicazione nel file system locale del nodo master nel cluster in Amazon S3. Puoi quindi inviare l'applicazione come flusso di lavoro batch.

In modalità batch, carica lo script Spark in Amazon S3 o nel file system del nodo master locale, quindi invia il lavoro al cluster come fase. Le fasi Spark possono essere inviate a un cluster di lunga durata o a un cluster transitorio.

Creazione di un cluster con Spark installato

Per avviare un cluster con Spark installato mediante la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Scegli Crea cluster.
3. Per Software Configuration (Configurazione software), scegliere la versione AMI necessaria.
4. Per Applications to be installed (Applicazioni da installare), scegliere Spark dall'elenco, quindi scegliere Configure and add (Configura e aggiungi).
5. Aggiungere argomenti per modificare la configurazione di Spark, come desiderato. Per ulteriori informazioni, consulta [Configurazione di Spark](#). Scegliere Add (Aggiungi).
6. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

L'esempio seguente mostra come creare un cluster con Spark utilizzando Java:

```
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
SupportedProductConfig sparkConfig = new SupportedProductConfig()
    .withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Spark Cluster")
    .withAmiVersion("3.11.0")
    .withNewSupportedProducts(sparkConfig)
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyName")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge")
    );
RunJobFlowResult result = emr.runJobFlow(request);
```

Configurazione di Spark

Puoi configurare Spark quando crei un cluster eseguendo l'operazione di bootstrap in [aws-labs/emr-bootstrap-actions/spark su Github](#). Per gli argomenti accettati dall'operazione di bootstrap, vedi il file [README](#) in quel repository. L'operazione di bootstrap configura proprietà nel file

`$SPARK_CONF_DIR/spark-defaults.conf`. Per ulteriori informazioni sulle impostazioni, consulta l'argomento relativo alla configurazione di Spark nella documentazione di Spark. Puoi sostituire "latest" nel seguente URL con il numero di versione di Spark da installare, ad esempio `2.2.0` <http://spark.apache.org/docs/latest/configuration.html>.

Puoi inoltre configurare Spark in modo dinamico al momento dell'invio di ogni applicazione. Un'impostazione che consente di massimizzare automaticamente l'allocazione delle risorse per un esecutore è disponibile nel file di configurazione `spark`. Per ulteriori informazioni, consulta [Sostituzione delle impostazioni di configurazione predefinite di Spark](#).

Modifica delle impostazioni di default di Spark

L'esempio seguente mostra come creare un cluster con `spark.executor.memory` impostato su 2G utilizzando AWS CLI.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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
```

Invio di lavoro a Spark

Per inviare lavoro a un cluster, utilizza una fase per eseguire lo script `spark-submit` nel cluster EMR. Aggiungi la fase utilizzando il metodo `addJobFlowSteps` in [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);
```

Sostituzione delle impostazioni di configurazione predefinite di Spark

I valori di configurazione predefiniti di Spark possono essere sostituiti per ogni applicazione. Puoi eseguire questa operazione quando invii applicazioni utilizzando una fase, che in pratica passa opzioni a `spark-submit`. Ad esempio, puoi cambiare la memoria assegnata a un processo esecutore modificando `spark.executor.memory`. Puoi fornire il parametro `--executor-memory` con un argomento simile al seguente:

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Nello stesso modo, puoi regolare `--executor-cores` e `--driver-memory`. In una fase, forniresti i seguenti argomenti alla fase:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

Puoi anche regolare impostazioni che possono non avere un parametro integrato utilizzando l'opzione `--conf`. Per ulteriori informazioni su altre impostazioni regolabili, consulta l'argomento [Dynamically Loading Spark Properties \(Caricamento dinamico delle proprietà Spark\)](#) nella documentazione di Apache Spark.

Differenze tra l'utility S3DistCp e le versioni AMI di Amazon EMR precedenti

Versioni di S3DistCp supportate in Amazon EMR

Le seguenti versioni di S3DistCp sono supportate nelle versioni AMI di Amazon EMR. Le versioni di S3DistCp successive alla versione 1.0.7 sono disponibili direttamente nei cluster. Utilizza il file JAR in `/home/hadoop/lib` per le caratteristiche più recenti.

Versione	Descrizione	Data di rilascio
1.0.8	Aggiunge le opzioni <code>--appendToLastFile</code> , <code>--requirePreviousManifest</code> e <code>--storageClass</code> .	3 gennaio 2014
1.0.7	Aggiunge l'opzione <code>--s3ServerSideEncryption</code> .	2 maggio 2013
1.0.6	Aggiunge l'opzione <code>--s3Endpoint</code> .	6 agosto 2012
1.0.5	Migliora la possibilità di specificare quale versione di S3DistCp eseguire.	27 giugno 2012
1.0.4	Migliora l'opzione <code>--deleteOnSuccess</code> .	19 giugno 2012
1.0.3	Aggiunge supporto per le opzioni <code>--numberOfFiles</code> e <code>--startingIndex</code> .	12 giugno 2012
1.0.2	Migliora la denominazione di file quando si utilizzano gruppi.	6 giugno 2012
1.0.1	Versione iniziale di S3DistCp.	19 gennaio 2012

Aggiunta di una fase di copia di S3DistCp a un cluster

Per aggiungere una fase di copia di S3DistCp a un cluster in esecuzione, digita il comando seguente, sostituisci `j-3GYXXXXXX9I0K` con l'ID del cluster e sostituisci `mybucket` con il nome di bucket Amazon S3.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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 Caricamento di log Amazon CloudFront in HDFS

Questo esempio carica i log di Amazon CloudFront in HDFS aggiungendo una fase a un cluster in esecuzione. Cambia inoltre il formato di compressione da Gzip (l'impostazione predefinita in CloudFront) in LZO. Ciò è utile in quanto i dati compressi con LZO possono essere suddivisi in più mappe non appena vengono decompressi, senza quindi attendere il completamento della compressione, come avviene con Gzip. Di conseguenza, si ha un miglioramento delle prestazioni durante l'analisi dei dati con Amazon EMR. L'esempio migliora le prestazioni anche mediante l'utilizzo dell'espressione regolare specificata nell'opzione `--groupBy` per combinare tutti i log di una determinata ora in un singolo file. I cluster Amazon EMR sono più efficienti quando si elaborano alcuni file voluminosi compressi con LZO anziché molti file di piccole dimensioni compressi con Gzip. Per suddividere i file LZO, è necessario indicizzarli e utilizzare la libreria di terza parte `hadoop-lzo`.

Per caricare i log di Amazon CloudFront in HDFS, digita il comando seguente, sostituisci `j-3GYXXXXXX9I0K` con l'ID del cluster e sostituisci `mybucket` con il nome di bucket Amazon S3.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```

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"]

```

Supponiamo che l'esempio precedente sia eseguito sui seguenti file di log di CloudFront.

```
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz
```

```
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3DistCp copi, concatena e comprime i file nei due file seguenti, dove il nome di file è determinato dalla corrispondenza stabilita dell'espressione regolare.

```
hdfs:///local/2012-02-23-01.lzo  
hdfs:///local/2012-02-23-02.lzo
```


Novità

Questa pagina descrive le modifiche e le funzionalità disponibili negli ultimi rilasci di Amazon EMR 6.x e Amazon EMR 5.x. Queste note di rilascio sono disponibili anche nella pagina [Versione 6.14.0 di Amazon EMR](#) e nella pagina [Versione 5.36.1 di Amazon EMR](#) insieme alle versioni dell'applicazione, alle versioni dei componenti e alle classificazioni di configurazione disponibili per ciascun rilascio.

Iscriviti al feed RSS per le note di rilascio di Amazon EMR all'indirizzo <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/amazon-emr-release-notes.rss> per ricevere aggiornamenti quando è disponibile un nuovo rilascio di Amazon EMR.

Per le note relative ai rilasci precedenti, consulta la sezione [Archivio delle note di rilascio di Amazon EMR](#).

Note

I rilasci di Amazon EMR ora utilizzano AWS Signature Version 4 (SigV4) per autenticare le richieste ad Amazon S3. Si consiglia di utilizzare un rilascio di Amazon EMR che supporti SigV4 in modo da poter continuare ad accedere ai nuovi bucket S3 ed evitare interruzioni dei carichi di lavoro. Per ulteriori informazioni e per un elenco dei rilasci di Amazon EMR che supportano SIGv4, consulta la sezione [Amazon EMR e AWS Signature Version 4](#).

Approccio per mitigare il CVE-2021-44228

Note

Per versione di rilascio 6.9.0 e successive di Amazon EMR, tutti i componenti installati da Amazon EMR che utilizzano librerie Log4j utilizzano Log4j versione 2.17.1 o successiva.

Amazon EMR eseguito su EC2

Il problema discusso nel [CVE-2021-44228](#) riguarda le versioni principali di Apache Log4j (dalla 2.0.0 alla 2.14.1) durante l'elaborazione di input da origini non attendibili. I cluster Amazon EMR avviati con i rilasci di Amazon EMR della serie 5.x fino al 5.34.0 e della serie 6.x fino al 6.5.0 includono framework open source come Apache Hive, Flink, HUDI, Presto e Trino, che utilizzano queste

versioni di Apache Log4j. Tuttavia, molti clienti utilizzano i framework open source installati sui cluster Amazon EMR per elaborare e registrare gli input da origini non attendibili.

Suggeriamo di applicare la "Soluzione per l'operazione di bootstrap di Amazon EMR relativa al CVE-2021-44228 di Log4j" come descritto nella sezione successiva. Questa soluzione si rivolge anche al CVE-2021-45046.

Note

Gli script dell'operazione bootstrap per Amazon EMR sono stati aggiornati il 7 settembre 2022 per includere correzioni di bug incrementali e miglioramenti per Oozie. Se utilizzi Oozie, devi applicare la soluzione dell'operazione di bootstrap di Amazon EMR aggiornata, descritta nella sezione seguente.

Amazon EMR su EKS

Se utilizzi [Amazon EMR su EKS](#) con la configurazione predefinita, non riscontrerai il problema descritto nel CVE-2021-44228 e non sarà necessario applicare la soluzione descritta nella sezione [Soluzione per l'operazione di bootstrap di Amazon EMR relativa al CVE-2021-44228 e al CVE-2021-45046 di Log4j](#). Per Amazon EMR su EKS, il runtime di Amazon EMR per Spark utilizza Apache Log4j versione 1.2.17. Se utilizzi Amazon EMR su EKS, evita di modificare le impostazioni predefinite per il componente `log4j . appender in log`.

Soluzione per l'operazione di bootstrap di Amazon EMR relativa al CVE-2021-44228 e al CVE-2021-45046 di Log4j

Questa soluzione fornisce un'operazione di bootstrap di Amazon EMR che deve essere applicata ai cluster Amazon EMR. Per ogni rilascio di Amazon EMR, troverai qui sotto un link a uno script di operazione di bootstrap. Per applicare questa operazione di bootstrap è necessario completare i seguenti passaggi:

1. Copia lo script che corrisponde al rilascio di Amazon EMR in un bucket S3 locale nel tuo Account AWS. Assicurati di utilizzare uno script di bootstrap specifico per il tuo rilascio di Amazon EMR.
2. Imposta un'operazione di bootstrap per i cluster EMR per eseguire lo script copiato nel bucket S3 secondo le istruzioni descritte in [Documentazione di EMR](#). Se sono state configurate altre operazioni di bootstrap per i cluster EMR, assicurati che questo script sia impostato come primo script di operazione di bootstrap da eseguire.

3. Termina i cluster EMR esistenti e avvia nuovi cluster con lo script di operazione di bootstrap.AWS consiglia di testare gli script di bootstrap nell'ambiente di test e di convalidare le applicazioni prima di applicarle all'ambiente di produzione. Se non si sta utilizzando la revisione più recente di una versione secondaria di EMR (ad esempio 6.3.0), è necessario utilizzare la revisione più recente (ad esempio, 6.3.1) e quindi applicare la soluzione di cui sopra.

CVE-2021-44228 e CVE-2021-45046 - Script di bootstrap per le release EMR

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
6.5.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.5.0-v2.sh</code>	24 marzo 2022
6.4.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh</code>	24 marzo 2022
6.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh</code>	24 marzo 2022
6.2.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh</code>	24 marzo 2022
6.1.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh</code>	14 dicembre 2021

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
6.0.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh</code>	14 dicembre 2021
5.34.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.34.0-v2.sh</code>	12 dicembre 2021
5.33.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.33.1-v2.sh</code>	12 dicembre 2021
5.32.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.32.1-v2.sh</code>	13 dicembre 2021
5.31.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.31.1-v2.sh</code>	13 dicembre 2021
5.30.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.30.2-v2.sh</code>	14 dicembre 2021

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
5.29.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.29.0-v2.sh</code>	14 dicembre 2021
5.28.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.28.1-v2.sh</code>	15 dicembre 2021
5.27.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</code>	15 dicembre 2021
5.26.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</code>	15 dicembre 2021
5.25.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</code>	15 dicembre 2021
5.24.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</code>	15 dicembre 2021

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
5.23.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</code>	15 dicembre 2021
5.22.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	15 dicembre 2021
5.21.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	15 dicembre 2021
5.20.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	15 dicembre 2021
5.19.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	15 dicembre 2021
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	15 dicembre 2021

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
5.17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	15 dicembre 2021
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	15 dicembre 2021
5.15.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	15 dicembre 2021
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	15 dicembre 2021
5.13.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</code>	15 dicembre 2021
5.12.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</code>	15 dicembre 2021

Versione del rilascio di Amazon EMR	Percorso script	Data di rilascio dello script
5.11.4	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</code>	15 dicembre 2021
5.10.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</code>	15 dicembre 2021
5.9.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh</code>	15 dicembre 2021
5.8.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh</code>	15 dicembre 2021
5.7.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh</code>	15 dicembre 2021

Versione del rilascio di EMR	Ultima revisione: dicembre 2021
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1

Versione del rilascio di EMR	Ultima revisione: dicembre 2021
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 o 5.30.1	5.30.2
5.28.0	5.28.1
5.27.0	5.27.1
5.24.0	5.24.1
5.23.0	5.23.1
5.21.0 o 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 o 5.17.1	5.17.2
5.16.0	5.16.1
5.15.0	5.15.1
5.14.0 o 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

Versione del rilascio di EMR	Ultima revisione: dicembre 2021
5.9.0	5.9.1
5.8.0, 5.8.1, 5.8.2	5.8.3
5.7.0	5.7.1

Domande frequenti

- Il CVE-2021-44228 interessa le versioni di EMR precedenti a EMR 5?

No. I rilasci di EMR precedenti al rilascio 5 di EMR utilizzano versioni Log4j precedenti alla 2.0.

- Questa soluzione si rivolge anche al CVE-2021-45046?

Sì, questa soluzione si rivolge anche al CVE-2021-45046.

- La soluzione gestisce le applicazioni personalizzate installate sui cluster EMR?

Lo script di bootstrap aggiorna solo i file JAR installati da EMR. Se installi ed esegui applicazioni personalizzate e file JAR sui cluster EMR tramite operazioni di bootstrap, come passaggi inviati ai cluster, utilizzando AMI Amazon Linux personalizzate o tramite qualsiasi altro meccanismo, mettiti in contatto con il fornitore dell'applicazione per determinare se le applicazioni personalizzate sono interessate dal CVE-2021-44228 e per trovare una soluzione appropriata.

- Come dovrei gestire le [immagini docker personalizzate](#) con EMR su EKS?

Se aggiungi applicazioni personalizzate ad Amazon EMR su EKS utilizzando [immagini docker personalizzate](#) oppure invii processi ad Amazon EMR su EKS con i file delle applicazioni personalizzati, collabora con il fornitore dell'applicazione per determinare se le tue applicazioni personalizzate sono interessate dal CVE-2021-44228 e per trovare una soluzione appropriata.

- Come funziona lo script bootstrap per mitigare il problema descritto in CVE-2021-44228 e CVE-2021-45046?

Lo script di bootstrap aggiorna le istruzioni di avvio di EMR aggiungendo un nuovo set di istruzioni. Queste nuove istruzioni eliminano i file di classe JndiLookup utilizzati tramite Log4j da tutti i framework open source installati da EMR. Questo segue il [suggerimento pubblicato da Apache](#) per risolvere i problemi di Log4j.

- Esiste un aggiornamento a EMR che utilizza Log4j versione 2.17.1 o successiva?

I rilasci di EMR dalla versione 5 fino alla 5.34 e dalla 6 fino alla 6.5 utilizzano versioni precedenti di framework open source incompatibili con le versioni più recenti di Log4j. Se si continua a utilizzare questo rilascio, consigliamo di applicare l'operazione di bootstrap per mitigare i problemi discussi nei CVE. Dopo i rilasci di EMR dalla versione 5 fino alla 5.34 e dalla 6 fino alla 6.5, le applicazioni che utilizzano Log4j 1.x e Log4j 2.x verranno aggiornate per utilizzare rispettivamente Log4j 1.2.17 (o successiva) e Log4j 2.17.1 (o successiva) e non richiederanno l'utilizzo delle operazioni di bootstrap sopra fornite per mitigare i problemi CVE.

- Il CVE-2021-45105 interessa le versioni di EMR?

Il CVE-2021-45105 non riguarda le applicazioni installate da Amazon EMR con le configurazioni di default di EMR. Tra le applicazioni installate da Amazon EMR, solo Apache Hive utilizza Apache Log4j con [ricerche di contesto](#); al contempo, non utilizza un layout di pattern non di default in un modo che consenta l'elaborazione di dati di input inappropriati.

- Amazon EMR è influenzato da una delle seguenti informazioni CVE?

La seguente tabella contiene un elenco di CVE correlati a Log4j e annota se ogni CVE influenza Amazon EMR. Le informazioni contenute in questa tabella si applicano solo quando le applicazioni sono installate da Amazon EMR utilizzando le configurazioni di default.

CVE	Impatti EMR	Note
CVE-2022-23302	No	Amazon EMR non configura Log4j JMSSink
CVE-2022-23305	No	Amazon EMR non configura Log4j JDBCAppender
CVE-2022-23307	No	Amazon EMR non configura Log4j Chainsaw
CVE-2020-9493	No	Amazon EMR non configura Log4j Chainsaw
CVE-2021-44832	No	Amazon EMR non configura Log4j DBCAppender con una stringa di connessione JNDI

CVE	Impatti EMR	Note
CVE-2021-4104	No	Amazon EMR non utilizza Log4j JMSAppender
CVE-2020-9488	No	Le applicazioni installate da Amazon EMR non utilizzano Log4j SMTPAppender
CVE-2019-17571	No	Amazon EMR blocca l'accesso pubblico ai cluster e non avvia SocketServer
CVE-2019-17531	No	Consigliamo di eseguire l'aggiornamento alla versione del rilascio di Amazon EMR più recente. Amazon EMR 5.33.0 e versioni successive utilizzano jackson-databind 2.6.7.4 o versioni successive e ed EMR 6.1.0 e versioni successive utilizzano jackson-databind 2.10.0 o versioni successive. Queste versioni di jackson-databind non sono influenzate dal CVE.

Amazon EMR 6.14.0 (versione più recente della serie 6.x)

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Le seguenti note di rilascio includono informazioni relative alla versione 6.14.0 di Amazon EMR. Le modifiche sono relative alla versione 6.13.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.
- La [scalabilità gestita di Amazon EMR](#) è ora disponibile nella regione ap-southeast-3 Asia Pacific (Giacarta) per i cluster creati con Amazon EMR 6.14.0 e versioni successive.

Modifiche, miglioramenti e problemi risolti

- La versione 6.14.0 ottimizza la gestione dei log con Amazon EMR in esecuzione su Amazon EC2. Di conseguenza, potresti notare una leggera riduzione dei costi di archiviazione per i log del cluster.
- La versione 6.14.0 migliora il flusso di lavoro di dimensionamento per tenere conto delle diverse istanze principali che presentano una variazione sostanziale delle dimensioni dei loro volumi Amazon EBS. Questo miglioramento si applica solo ai nodi principali; le operazioni di riduzione delle dimensioni per i nodi di attività non sono influenzate.
- La versione 6.14.0 migliora il modo in cui Amazon EMR interagisce con applicazioni open source come Apache Hadoop YARN ResourceManager and HDFS NameNode. Questo miglioramento riduce il rischio di ritardi operativi dovuti nel dimensionamento del cluster e mitiga gli errori di avvio dovuti a problemi di connettività con le applicazioni open source.
- La versione 6.14.0 ottimizza l'installazione delle applicazioni all'avvio del cluster. Ciò migliora i tempi di avvio del cluster per determinate combinazioni di applicazioni Amazon EMR.
- La versione 6.14.0 corregge un problema a causa del quale le operazioni di riduzione del cluster potrebbero bloccarsi quando un cluster in esecuzione in un VPC con un dominio personalizzato rileva un riavvio del nodo principale o del nodo attività.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230906.0	4.14.322	11 settembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

Amazon EMR 5.36.1 (versione più recente della serie 5.x)

I nuovi rilasci di Amazon EMR diventano disponibili in Regioni differenti nell'arco di alcuni giorni, a partire dalla prima Regione alla data del rilascio iniziale. La versione più recente potrebbe non essere disponibile nella tua Regione durante questo periodo.

Le seguenti note di rilascio includono informazioni relative alla versione 5.36.1 di Amazon EMR. Le modifiche sono relative alla versione 5.36.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- La versione 5.36.1 di Amazon EMR aggiunge il supporto per l'archiviazione dei log su Amazon S3 durante la riduzione dei cluster. Nelle precedenti versioni 5.x, era possibile archiviare i file di log in Amazon S3 solo durante la terminazione del cluster. Questo miglioramento garantisce la persistenza su Amazon S3 dei file di log generati nel cluster anche dopo la terminazione del nodo. Per ulteriori informazioni, consulta [Configurazione della registrazione e del debug di cluster](#).
- La versione 5.36.1 migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- La versione 5.36.1 riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- La versione 5.36.1 corregge un problema per cui i daemon Amazon EMR sul nodo primario mantenevano i metadati obsoleti per le istanze terminate nel cluster. Il mantenimento di dati

obsoleti potrebbe causare un aumento illimitato dell'utilizzo della CPU e della memoria all'interno del cluster e, infine, provocare errori del cluster.

- Per i cluster avviati con più nodi primari, la versione 5.36.1 corregge un problema per cui un guasto hardware di Amazon EC2 su uno dei nodi primari poteva causare l'esito negativo di un secondo nodo primario e rendere instabile il cluster.
- Per i cluster configurati con la crittografia in transito, ora Scalabilità gestita riconosce i dati shuffle di Spark. I dati shuffle di Spark sono dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche. Durante la riduzione, Scalabilità gestita ignora le istanze con dati shuffle. In questo modo si evitano nuovi tentativi di eseguire il processo e nuovi calcoli, che sono costosi in termini di prezzo e prestazioni. Per ulteriori informazioni sulle operazioni di shuffle, consulta la [Guida di programmazione Spark](#).
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	(Centrale), Israele (Tel Aviv) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Amazon EMR e AWS Signature Version 4

I rilasci di Amazon EMR ora utilizzano AWS Signature Version 4 (SigV4) per autenticare le richieste ad Amazon S3. I bucket creati in Amazon S3 dopo il 24 giugno 2020 non supportano le richieste

firmate da Signature Version 2 (SigV2). I bucket creati fino al 24 giugno 2020 incluso continueranno a supportare SigV2. Si consiglia di eseguire la migrazione a un rilascio di Amazon EMR che supporti SigV4 in modo da poter continuare ad accedere ai nuovi bucket S3 ed evitare interruzioni dei carichi di lavoro.

Se utilizzi applicazioni incluse in Amazon EMR, come Apache Spark, Apache Hive, Presto e così via, non è necessario modificare il codice dell'applicazione per utilizzare SigV4. Se utilizzi applicazioni personalizzate che non sono incluse in Amazon EMR, potrebbe essere necessario aggiornare il codice per utilizzare SigV4. Per ulteriori informazioni, consulta la sezione [Moving from Signature Version 2 to Signature Version 4](#) (Passaggio da Signature Version 2 a Signature Version 4) nella Guida per l'utente di Amazon S3.

I seguenti rilasci di Amazon EMR supportano 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 e emr-5.22.0 e successivi.

Archivio delle note di rilascio di Amazon EMR

Le note di rilascio per tutti i rilasci di Amazon EMR sono disponibili di seguito. Per informazioni complete su ogni rilascio, consulta le sezioni [Versioni di rilascio di Amazon EMR 6.x](#), [Versioni del rilascio di Amazon EMR 5.x](#) e [Versioni del rilascio di Amazon EMR 4.x](#).

Iscriviti al feed RSS per le note di rilascio di Amazon EMR all'indirizzo <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/amazon-emr-release-notes.rss> per ricevere aggiornamenti quando è disponibile un nuovo rilascio di Amazon EMR.

Versione 6.14.0

Le seguenti note di rilascio includono informazioni relative alla versione 6.14.0 di Amazon EMR. Le modifiche sono relative alla versione 6.13.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.
- La [scalabilità gestita di Amazon EMR](#) è ora disponibile nella regione ap-southeast-3 Asia Pacific (Giacarta) per i cluster creati con Amazon EMR 6.14.0 e versioni successive.

Modifiche, miglioramenti e problemi risolti

- La versione 6.14.0 ottimizza la gestione dei log con Amazon EMR in esecuzione su Amazon EC2. Di conseguenza, potresti notare una leggera riduzione dei costi di archiviazione per i log del cluster.
- La versione 6.14.0 migliora il flusso di lavoro di dimensionamento per tenere conto delle diverse istanze principali che presentano una variazione sostanziale delle dimensioni dei loro volumi Amazon EBS. Questo miglioramento si applica solo ai nodi principali; le operazioni di riduzione delle dimensioni per i nodi di attività non sono influenzate.
- La versione 6.14.0 migliora il modo in cui Amazon EMR interagisce con applicazioni open source come Apache Hadoop YARN ResourceManager and HDFS NameNode. Questo miglioramento riduce il rischio di ritardi operativi dovuti nel dimensionamento del cluster e mitiga gli errori di avvio dovuti a problemi di connettività con le applicazioni open source.
- La versione 6.14.0 ottimizza l'installazione delle applicazioni all'avvio del cluster. Ciò migliora i tempi di avvio del cluster per determinate combinazioni di applicazioni Amazon EMR.
- La versione 6.14.0 corregge un problema a causa del quale le operazioni di riduzione del cluster potrebbero bloccarsi quando un cluster in esecuzione in un VPC con un dominio personalizzato rileva un riavvio del nodo principale o del nodo attività.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023.906.0	4.14.322	11 settembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

Versione 6.13.0

Le seguenti note di rilascio includono informazioni relative alla versione 6.13.0 di Amazon EMR. Le modifiche sono relative alla versione 6.12.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.

Modifiche, miglioramenti e problemi risolti

- La versione 6.13.0 migliora il daemon di gestione dei log di Amazon EMR per garantire che tutti i log vengano caricati a cadenza regolare su Amazon S3 quando viene emesso un comando di terminazione del cluster. Ciò facilita le terminazioni più rapide dei cluster.
- La versione 6.13.0 migliora le funzionalità di gestione dei log di Amazon EMR per garantire il caricamento coerente e tempestivo di tutti i file di log su Amazon S3. Ciò avvantaggia in particolare i cluster EMR di lunga durata.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-808.0	4.14.320	24 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

Versione 6.12.0

Le seguenti note di rilascio includono informazioni relative alla versione 6.12.0 di Amazon EMR. Le modifiche sono relative alla versione 6.11.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Nuove funzionalità di

- 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.
- Le versioni 6.12.0 e successive di Amazon EMR supportano l'integrazione LDAP con Apache Livy, Apache Hive tramite HiveServer2 (HS2), Trino, Presto e Hue. Inoltre, puoi installare Apache Spark e Apache Hadoop su un cluster EMR che utilizza 6.12.0 o versioni successive e configurarli per l'utilizzo di LDAP. Per ulteriori informazioni, consulta [Utilizzo dei server Active Directory o LDAP per l'autenticazione con Amazon EMR](#).

Modifiche, miglioramenti e problemi risolti

- Le versioni 6.12.0 e successive di Amazon EMR forniscono il supporto di runtime Java 11 per Flink. Per ulteriori informazioni, consulta [Configurazione di Flink per l'esecuzione con Java 11](#).
- Amazon EMR 6.12.0 supporta tutte le applicazioni con Amazon Corretto 8 per impostazione predefinita, ad eccezione di Trino. Per Trino, Amazon EMR supporta Amazon Corretto 17 per impostazione predefinita a partire dalla versione 6.9.0 di Amazon EMR. Amazon EMR supporta anche alcune applicazioni con Amazon Corretto 11 e 17. Queste applicazioni sono elencate nella seguente tabella. Se desideri modificare la JVM predefinita sul cluster, segui le istruzioni riportate in [Configurazione delle applicazioni per l'uso di una macchina virtuale Java specifica](#) per ogni applicazione in esecuzione sul cluster. Puoi utilizzare solo una delle versioni di runtime Java per cluster. Amazon EMR non supporta l'esecuzione di nodi o applicazioni differenti su versioni di runtime diverse sullo stesso cluster.

Sebbene Amazon EMR supporti Amazon Corretto 11 e 17 su Apache Spark, Apache Hadoop e Apache Hive, le prestazioni potrebbero peggiorare per alcuni carichi di lavoro quando utilizzi queste versioni di Corretto. Consigliamo di testare i carichi di lavoro prima di modificare i valori predefiniti.

Versioni Java predefinite per le applicazioni in Amazon EMR 6.12

Applicazione	Versione Java/Amazon Corretto (l'impostazione predefinita è in grassetto)
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
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8

Applicazione	Versione Java/Amazon Corretto (l'impostazione predefinita è in grassetto)
Pig	8
Zookeeper	8

- La versione 6.12.0 aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.
- La versione 6.12.0 corregge un problema a causa del quale le operazioni di riduzione del cluster potrebbero bloccarsi quando un nodo principale sottoposto a una disattivazione graduale perde integrità per qualsiasi motivo prima della completa disattivazione.
- La versione 6.12.0 migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- La versione 6.12.0 migliora le prestazioni e l'efficienza del servizio di monitoraggio dell'integrità per Amazon EMR aumentando la velocità con cui registra le modifiche di stato per le istanze. Questo miglioramento riduce la possibilità di un peggioramento delle prestazioni per i nodi del cluster che eseguono più strumenti client personalizzati o applicazioni di terze parti.
- La versione 6.12.0 migliora le prestazioni del daemon di gestione dei log sul cluster per Amazon EMR. Di conseguenza, ci sono meno possibilità di ridurre le prestazioni con i cluster EMR che eseguono fasi con elevata simultaneità.
- Con la versione 6.12.0 di Amazon EMR, il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- La versione 6.12.0 include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.

- La versione 6.12.0 consente la rotazione dei log di YARN Timeline Server. Ciò riduce al minimo gli scenari di utilizzo eccessivo del disco, in particolare per i cluster di lunga durata.
- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-0727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

Versione 6.11.1

Le seguenti note di rilascio includono informazioni relative alla versione 6.11.1 di Amazon EMR. Le modifiche sono relative alla versione 6.11.0. Per ulteriori informazioni sulla tempistica di pubblicazione, consulta il [Log delle modifiche](#).

Modifiche, miglioramenti e problemi risolti

- A causa di un conflitto di blocco, un nodo può entrare in una situazione di stallo se viene aggiunto o rimosso contemporaneamente al tentativo di disattivazione. Di conseguenza, Hadoop Resource Manager (YARN) non risponde e influisce su tutti i container in entrata e quelli attualmente in esecuzione.
- Questa versione include una modifica che consente ai cluster ad alta disponibilità di ripristinare lo stato di errore dopo il riavvio.
- Questa versione include correzioni di sicurezza per Hue e HBase.
- Questa versione corregge un problema per cui i cluster che eseguono carichi di lavoro su Spark con Amazon EMR potrebbero ricevere silenziosamente risultati errati con `contains`, `startsWith`, `endsWith` e `like`. Questo problema si verifica quando utilizzi le espressioni su campi partizionati che contengono metadati in Amazon EMR Hive3 Metastore Server (HMS).
- Questa versione corregge un problema di limitazione (della larghezza di banda della rete) sul lato Glue quando non sono presenti funzioni definite dall'utente (UDF).
- Questa versione corregge un problema che comporta l'eliminazione dei log dei container tramite il servizio di aggregazione dei log dei nodi prima che il pusher possa inviarli a S3 in caso di disattivazione di YARN.
- Questa versione corregge un problema con i parametri di FairShare Scheduler quando Node Label è abilitato per Hadoop.
- Questa versione corregge un problema che influiva sulle prestazioni di Spark quando veniva impostato un valore `true` predefinito per la configurazione di `spark.yarn.heterogeneousExecutors.enabled` in `spark-defaults.conf`.
- Questa versione corregge un problema a causa del quale Reduce Task non era in grado di leggere i dati shuffle. Il problema ha causato errori nelle query di Hive con un errore di memoria danneggiata.
- Questa versione aggiunge un nuovo meccanismo di tentativi al flusso di lavoro dimensionabile per i cluster EMR che eseguono Presto o Trino. Questo miglioramento riduce la possibilità che il ridimensionamento del cluster si blocchi a tempo indeterminato a causa di una singola operazione

di ridimensionamento con esito negativo. Inoltre, migliora l'utilizzo del cluster, poiché quest'ultimo aumenta e riduce più rapidamente.

- Questa versione migliora la logica di riduzione del cluster in modo che quest'ultimo non tenti di ridurre i nodi principali al di sotto dell'impostazione del fattore di replica HDFS per il cluster. Ciò è in linea con i requisiti di ridondanza dei dati e riduce la possibilità che un'operazione di dimensionamento si blocchi.
- Il daemon di gestione dei log è stato aggiornato per identificare tutti i log che sono in uso attivo con handle di file aperti sull'archiviazione dell'istanza locale e i processi associati. Questo aggiornamento garantisce che Amazon EMR elimini correttamente i file e recuperi lo spazio di archiviazione dopo la memorizzazione dei log su Amazon S3.
- Questa versione include un miglioramento del daemon di gestione dei log che elimina le directory delle fasi vuote e inutilizzate nel file system del cluster locale. Un numero eccessivo di directory vuote può ridurre le prestazioni dei daemon Amazon EMR e causare un utilizzo eccessivo del disco.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023.727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Versione 6.11.0

Le seguenti note di rilascio includono informazioni relative alla versione 6.11.0 di Amazon EMR. Le modifiche sono relative alla versione 6.10.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove funzionalità di

- Amazon EMR 6.11.0 supporta 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 e PrestoDB 0.279-amzn-0.

Modifiche, miglioramenti e problemi risolti

- Con Amazon EMR 6.11.0, il connettore DynamoDB è stato aggiornato alla versione 5.0.0. La versione 5.0.0 utilizza AWS SDK for Java 2.x. Le versioni precedenti utilizzavano AWS SDK for Java 1.x. Come risultato di questo aggiornamento, consigliamo vivamente di testare il codice prima di utilizzare il connettore DynamoDB con Amazon EMR 6.11.
- Quando il connettore DynamoDB per Amazon EMR 6.11.0 chiama il servizio DynamoDB, utilizza il valore della regione fornito per la proprietà di `dynamodb.endpoint`. Consigliamo di configurare anche `dynamodb.region` quando utilizzi `dynamodb.endpoint` e che entrambe le proprietà abbiano come destinazione la stessa Regione AWS. Se utilizzi `dynamodb.endpoint` e non configuri `dynamodb.region`, il connettore DynamoDB per Amazon EMR 6.11.0 restituisce un'eccezione della regione non valida e tenta di riconciliare le informazioni della Regione AWS dal servizio di metadati di istanza (IMDS) Amazon EC2. Se il connettore non è in grado di recuperare la regione da IMDS, il valore predefinito è Stati Uniti orientali (Virginia settentrionale) (`us-east-1`). L'errore seguente è un esempio dell'eccezione della regione non valida che potresti ottenere se non configuri correttamente la proprietà `dynamodb.region`: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` per ulteriori informazioni sulle classi interessate dall'aggiornamento di AWS SDK for Java a 2.x, consulta il commit [Aggiornamento di AWS SDK for Java da 1.x a 2.x \(#175\)](#) nel repository GitHub per il connettore Amazon EMR - DynamoDB.
- Questa versione corregge un problema per cui i dati diventano NULL quando utilizzi Delta Lake per archiviare i dati delle tabelle Delta in Amazon S3 dopo l'operazione di ridenominazione delle colonne. Per ulteriori informazioni su questa funzionalità sperimentale di Delta Lake, consulta [Operazione di ridenominazione delle colonne](#) nella Guida per l'utente di Delta Lake.
- La versione 6.11.0 corregge un problema che potrebbe verificarsi quando si crea un nodo edge replicando uno dei nodi primari da un cluster con più nodi primari. Il nodo edge replicato potrebbe causare ritardi nelle operazioni di riduzione o comportare un elevato utilizzo della memoria sui nodi primari. Per ulteriori informazioni su come creare un nodo edge per comunicare con il cluster EMR, consulta [Strumento per creare nodi edge](#) nel repository `aws-samples` su GitHub.
- La versione 6.11.0 migliora il processo di automazione utilizzato da Amazon EMR per rimontare i volumi Amazon EBS su un'istanza dopo un riavvio.
- La versione 6.11.0 corregge un problema che causava discontinuità nei parametri Hadoop pubblicati da Amazon EMR su Amazon CloudWatch.
- La versione 6.11.0 corregge un problema con i cluster EMR per cui un aggiornamento del file di configurazione YARN che contiene l'elenco di esclusione dei nodi per il cluster viene interrotto a

causa dell'eccessivo utilizzo del disco. L'aggiornamento incompleto ostacola le future operazioni di riduzione del cluster. Questa versione garantisce l'integrità del cluster e il corretto funzionamento delle operazioni di dimensionamento.

- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con le versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` è impostato su `false` in `yarn-site.xml` per risolvere questo problema.

Sebbene la correzione si occupi dei problemi introdotti da YARN-9608, potrebbe influire sulla corretta esecuzione dei processi di Hive a causa della perdita dei dati shuffle sui cluster con Scalabilità gestita abilitata. In questa versione abbiamo mitigato tale rischio impostando anche `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` per i carichi di lavoro di Hive. Questa configurazione è disponibile solo con Amazon EMR 6.11.0 e versioni successive.

- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

Versione 6.10.0

Le seguenti note di rilascio includono informazioni relative alla versione 6.10.0 di Amazon EMR. Le modifiche sono relative alla versione 6.9.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove funzionalità di

- Amazon EMR 6.10.0 supporta Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 e PrestoDB 0.278.1.
- Amazon EMR 6.10.0 include un connettore Trino-Hudi nativo che fornisce l'accesso in lettura ai dati nelle tabelle Hudi. Puoi attivare il connettore con `trino-cli --catalog hudi` e configurarlo in base alle tue esigenze con `trino-connector-hudi`. L'integrazione nativa con Amazon EMR indica che non è più necessario utilizzare `trino-connector-hive` per eseguire query sulle tabelle Hudi. Per un elenco delle configurazioni supportate con il nuovo connettore, consulta la pagina [Connettori Hudi](#) nella documentazione di Trino.
- Le versioni 6.10.0 e successive di Amazon EMR supportano l'integrazione di Apache Zeppelin con Apache Flink. Per ulteriori informazioni, consulta [Utilizzo dei processi Flink di Zeppelin in Amazon EMR](#).

Problemi noti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. Questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Per risolvere questo problema in Amazon EMR 6.10.0, puoi impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Amazon EMR 6.10.0 rimuove la dipendenza da `minimal-json.jar` per l'[integrazione Amazon Redshift per Apache Spark](#) e aggiunge automaticamente i jar necessari relativi a Spark-Redshift al percorso della classe dell'esecutore per Spark: `spark-redshift.jar`, `spark-avro.jar` e `RedshiftJDBC.jar`.
- La versione 6.10.0 migliora il daemon di gestione dei log sul cluster per monitorare cartelle di log aggiuntive nel cluster EMR. Questo miglioramento riduce al minimo gli scenari di utilizzo eccessivo del disco.
- La versione 6.10.0 riavvia automaticamente il daemon di gestione dei log sul cluster quando si arresta. Questo miglioramento riduce la possibilità che i nodi appaiano non integri a causa dell'eccessivo utilizzo del disco.
- Amazon EMR 6.10.0 supporta gli endpoint regionali per la mappatura degli utenti EMRFS.
- La dimensione predefinita del volume root è aumentata a 15 GB in Amazon EMR 6.10.0 e versioni successive. Le versioni precedenti avevano una dimensione predefinita del volume root di 10 GB.
- La versione 6.10.0 corregge un problema che causava il blocco dei processi Spark quando tutti gli esecutori Spark rimanenti si trovavano su un host in fase di disattivazione con il gestore di risorse YARN.
- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione

della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-0727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			(Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

Rilascio 6.9.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.9.0. Le modifiche sono relative al rilascio di Amazon EMR 6.8.0. Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Nuove caratteristiche

- Amazon EMR rilascio 6.9.0 supporta Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 e Tez 0.10.2.
- Amazon EMR rilascio 6.9.0 include una nuova applicazione open source, [Delta Lake](#) 2.1.0.
- L'integrazione di Amazon Redshift per Apache Spark è inclusa in Amazon EMR rilascio 6.9.0 e successivi. In precedenza uno strumento open source, l'integrazione nativa è un connettore Spark che è possibile utilizzare per creare applicazioni Apache Spark in grado di leggere e scrivere dati in Amazon Redshift e Amazon Redshift Serverless. Per ulteriori informazioni, consulta [Utilizzo dell'integrazione di Amazon Redshift per Apache Spark su Amazon EMR](#).
- Amazon EMR rilascio 6.9.0 aggiunge il supporto per l'archiviazione dei log su Amazon S3 durante la riduzione dei cluster. In precedenza, era possibile archiviare i file di log in Amazon S3 solo durante la terminazione del cluster. La nuova funzionalità garantisce la persistenza su Amazon S3 dei file di log generati nel cluster anche dopo la terminazione del nodo. Per ulteriori informazioni, consulta [Configurazione della registrazione e del debug di cluster](#).
- Per supportare le query di lunga durata, Trino ora include un meccanismo di esecuzione con tolleranza ai guasti. L'esecuzione con tolleranza ai guasti mitiga gli errori delle query ritentando le query non riuscite o le attività dei loro componenti. Per ulteriori informazioni, consulta [Esecuzione con tolleranza ai guasti in Trino](#).
- È possibile utilizzare Apache Flink su Amazon EMR per l'elaborazione BATCH e STREAM unificata delle tabelle Apache Hive o dei metadati di qualsiasi origine di tabella Flink, come Iceberg, Kinesis o Kafka. È possibile specificare AWS Glue Data Catalog come metastore per Flink utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Per ulteriori informazioni, consulta [Configurazione di Flink in Amazon EMR](#).
- Ora è possibile specificare i ruoli di runtime AWS Identity and Access Management (IAM) e il controllo degli accessi basato su AWS Lake Formation per Apache Spark, Apache Hive e Presto su Amazon EMR sui cluster EC2 con Amazon SageMaker Studio. Per ulteriori informazioni, consulta [Configurazione dei ruoli di runtime per le fasi di Amazon EMR](#).

Problemi noti

- Per Amazon EMR rilascio 6.9.0, Trino non funziona su cluster abilitati per Apache Ranger. Se hai la necessità di utilizzare Trino con Ranger, contatta il [AWS Support](#).
- Se utilizzi l'integrazione di Amazon Redshift per Apache Spark e disponi di un'indicazione temporale `time`, `timez`, `timestamp` o `timestampz` con una precisione di microsecondi in formato Parquet, il connettore arrotonda i valori temporali al valore in millisecondi più vicino. Come soluzione alternativa, utilizza il parametro `unload_s3_format` del formato di scaricamento del testo.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Le connessioni ai cluster Amazon EMR da Amazon SageMaker Studio possono fallire in modo intermittente con un codice di risposta 403 Forbidden. Questo errore si verifica quando l'impostazione del ruolo IAM sul cluster richiede più di 60 secondi. Come soluzione alternativa, puoi installare una patch Amazon EMR per consentire nuovi tentativi e aumentare il timeout a un minimo di 300 secondi. Completa la seguente procedura per applicare l'operazione bootstrap all'avvio del cluster.

1. Scarica lo script di bootstrap e i file RPM tramite i seguenti URL Amazon S3.

```
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. Carica i file della fase precedente in un bucket Amazon S3 di tua proprietà. Il bucket deve trovarsi nella stessa Regione AWS in cui intendi avviare il cluster.
3. Includi la seguente operazione di bootstrap durante l'avvio del cluster EMR. Sostituisci *bootstrap_URI* e *RPM_URI* con gli URI corrispondenti di Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi SecretAgent e RecordServer servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà Log4j2. A causa della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink fornisce connettori nativi per i file system S3 e Hadoop che consentono alle applicazioni di creare un FileSink e di scrivere i dati in Amazon S3. Questo FileSink fallisce con una delle due eccezioni seguenti.

```
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:?]

```

Come soluzione alternativa, puoi installare una patch Amazon EMR, che risolve il problema sopra riportato in Flink. Per applicare l'operazione bootstrap all'avvio del cluster, completa la seguente procedura.

1. Scarica il [flink-rpm](#) nel bucket Amazon S3. Il tuo percorso RPM è `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Scarica lo script di bootstrap e i file RPM da Amazon S3 utilizzando il seguente URI. Sostituisci *regionName* con la Regione AWS in cui prevedi di avviare il cluster.

```

s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh

```

3. Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. In Amazon EMR 6.8.0 e 6.9.0, questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con [Amazon EMR 6.10.0](#), esiste una soluzione alternativa a questo problema per impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Per la release 6.9.0 e successive di Amazon EMR, tutti i componenti installati da Amazon EMR che utilizzano librerie Log4j utilizzano Log4j versione 2.17.1 o successiva.
- Quando utilizzi il connettore DynamoDB con Spark nelle versioni 6.6.0, 6.7.0 e 6.8.0 di Amazon EMR, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Amazon EMR rilascio 6.9.0 risolve questo problema.

- Amazon EMR 6.9.0 aggiunge un supporto limitato per il controllo degli accessi basato su Lake Formation con Apache Hudi durante la lettura dei dati tramite Spark SQL. Il supporto è per le query SELECT che utilizzano Spark SQL ed è limitato al controllo degli accessi a livello di colonna. Per ulteriori informazioni, consulta [Hudi e Formation](#).
- Quando usi Amazon EMR 6.9.0 per creare un cluster Hadoop con le [Etichette nodo](#) abilitate, l'[API delle metriche YARN](#) restituisce informazioni aggregate su tutte le partizioni, anziché sulla partizione predefinita. Per ulteriori informazioni, consulta [YARN-11414](#).
- Con Amazon EMR 6.9.0, abbiamo aggiornato Trino alla versione 398, che utilizza Java 17. La versione precedente supportata di Trino per Amazon EMR 6.8.0 era Trino 388 in esecuzione su Java 11. Per ulteriori informazioni su questa modifica, consulta [Aggiornamenti di Trino a Java 17](#) sul blog di Trino.
- Questa versione corregge un problema di mancata corrispondenza della sequenza temporale tra Apache BigTop e Amazon EMR nella sequenza di avvio del cluster EC2. Questa mancata corrispondenza della sequenza temporale si verifica quando un sistema tenta di eseguire due o più operazioni contemporaneamente anziché eseguirle nella sequenza corretta. Di conseguenza, alcune configurazioni del cluster hanno registrato timeout di avvio delle istanze e tempi di avvio del cluster più lenti.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Arabi Uniti), Canada (Centrale), Israele (Tel Aviv)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	(Centrale), Israele (Tel Aviv) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	12 gennaio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Rilascio 6.8.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.8.0. Le modifiche sono relative alla versione 6.7.0

Nuove caratteristiche

- Le fasi di Amazon EMR supportano gli endpoint Apache Livy e i client JDBC/ODBC. Per ulteriori informazioni, consulta [Configurazione dei ruoli di runtime per le fasi di Amazon EMR](#).
- Il rilascio 6.8.0 di Amazon EMR include Spark per Apache HBase 2.4.12. Con questa versione di HBase, è possibile archiviare ed eliminare le tabelle HBase. Il processo di archiviazione di Amazon S3 rinomina tutti i file della tabella nella directory di archivio. Questo processo può essere lungo e costoso. Ora è possibile saltare il processo di archiviazione ed eliminare rapidamente tabelle di grandi dimensioni. Per ulteriori informazioni, consulta [Utilizzo della shell HBase](#).

Problemi noti

- Hadoop 3.3.3 ha introdotto una modifica in YARN ([YARN-9608](#)) che mantiene i nodi in cui erano in esecuzione i container in uno stato di disattivazione fino al completamento dell'applicazione. Questa modifica assicura che i dati locali, come i dati shuffle, non vadano persi e non è necessario eseguire nuovamente il processo. In Amazon EMR 6.8.0 e 6.9.0, questo approccio potrebbe anche portare a un utilizzo insufficiente delle risorse sui cluster con o senza Scalabilità gestita abilitata.

Con [Amazon EMR 6.10.0](#), esiste una soluzione alternativa a questo problema per impostare il valore di `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` su `false` in `yarn-site.xml`. Nelle versioni 6.11.0 e successive di Amazon EMR, nonché 6.8.1, 6.9.1 e 6.10.1, la configurazione è impostata su `false` per impostazione predefinita per risolvere questo problema.

Modifiche, miglioramenti e problemi risolti

- Quando i rilasci 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggevano le tabelle di Apache Phoenix tramite la shell Apache Spark, Amazon EMR produceva un `NoSuchMethodError`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- Il rilascio 6.8.0 di Amazon EMR include [Apache Hudi](#) 0.11.1; tuttavia, i cluster Amazon EMR 6.8.0 sono compatibili anche con l'open source `hudi-spark3.3-bundle_2.12` di Hudi 0.12.0.
- Il rilascio 6.8.0 di Amazon EMR con Apache Spark 3.3.0. Questa versione di Spark utilizza Apache Log4j 2 e il file `log4j2.properties` per configurare Log4j nei processi Spark. Se utilizzi Spark nel cluster o crei cluster EMR con parametri di configurazione personalizzati e desideri eseguire l'aggiornamento alla versione 6.8.0 di Amazon EMR, devi migrare alla nuova classificazione della

configurazione e formato chiave `spark-log4j2` per Apache Log4j 2. Per ulteriori informazioni, consulta [Migrazione da Apache Log4j 1.x a Log4j 2.x](#).

- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

Note

Questa versione non riceve più aggiornamenti automatici delle AMI poiché è stata sostituita da un'altra versione di patch. La versione della patch è indicata dal numero dopo il secondo separatore decimale (6.8.**1**). Per scoprire se stai utilizzando l'ultima versione della patch, controlla le versioni disponibili nella [Guida alle versioni](#) o consulta il menu a discesa Versione Amazon EMR quando crei un cluster nella console o utilizzi l'API [ListReleaseLabels](#) o l'operazione [list-release-labels](#) nella CLI. Per ricevere aggiornamenti sulle nuove versioni, iscriviti al feed RSS nella pagina [Novità](#).

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Asia Pacifico (Melbourne), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 2023	Arabi Uniti), Canada (Centrale) Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022.0.912.1	4.14.291	6 settembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Problemi noti

- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la

divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. Ca sua della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Per ulteriori informazioni sulla tempistica di rilascio, consulta il [log delle modifiche](#).

Rilascio 6.7.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.7.0. Le modifiche sono relative alla versione 6.6.0.

Data del rilascio iniziale: 15 luglio 2022

Nuove caratteristiche

- Amazon EMR ora supporta Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 e Trino 0.378.
- Supporta i controlli degli accessi basati su ruoli IAM e Lake Formation con fasi EMR (Spark, Hive) per Amazon EMR sui cluster EC2.
- Supporta le istruzioni di definizione dei dati Apache Spark su cluster abilitati Apache Ranger. Adesso, include il supporto per le applicazioni Trino che leggono e scrivono metadati Apache Hive su cluster abilitati Apache Ranger. Per ulteriori informazioni, consulta la sezione [Abilitazione della governance federata utilizzando Trino e Apache Ranger su Amazon EMR](#).
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale),

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 719.0	4.14.287	10 agosto 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022606.1	4.14.281	15 luglio 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché

Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.

- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. A causa della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Rilascio 6.6.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.6.0. Le modifiche sono relative alla versione 6.5.0.

Data del rilascio iniziale: 9 maggio 2022

Ultimo aggiornamento della documentazione: 15 giugno 2022

Nuove caratteristiche

- Amazon EMR 6.6 ora supporta Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 e PrestoDB 0.267.
- Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Per ulteriori informazioni, consulta [Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR](#).

OsRelease Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023-727.0	4.14.320	14 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			li (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023719.0	4.14.320	2 agosto 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Spagna), Europa (Francoforte), Europa (Zurigo), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Hyderabad), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Medio Oriente (Emirati Arabi Uniti), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230628.0	4.14.318	12 luglio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230612.0	4.14.314	23 giugno 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202307.0	4.14.304	22 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023119.1	4.14.301	3 febbraio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.202210.1	4.14.301	22 dicembre 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022103.3	4.14.296	5 dicembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2022 805.0	4.14.287	30 agosto 2022	us-west-1

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022719.0	4.14.287	10 agosto 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022.0426.0	4.14.281	10 giugno 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsReleaseLabel (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022406.1	4.14.275	2 maggio 2022	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapore), Asia Pacifico (Sydney), Asia Pacifico (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

- Le applicazioni Amazon EMR versione 6.6 e successive che usano Log4j 1.x e Log4j 2.x sono aggiornate per usare rispettivamente Log4j 1.2.17 (o superiore) e Log4j 2.17.1 (o superiore) e non richiedono l'uso di [operazioni di bootstrap](#) per mitigare i problemi CVE.
- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento

gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).

- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.

Modifiche, miglioramenti e problemi risolti

- Amazon EMR riduce in media i tempi di avvio del cluster fino a 80 secondi per i cluster che utilizzano l'opzione AMI di default EMR e installano solo applicazioni comuni, come Apache Hadoop, Apache Spark e Apache Hive.

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- Quando usi il connettore DynamoDB con Spark nelle versioni di Amazon EMR 6.6.0, 6.7.0 e 6.8.0, tutte le letture della tabella restituiscono un risultato vuoto, anche se la divisione di input fa riferimento a dati non vuoti. Questo perché Spark 3.2.0 imposta `spark.hadoopRDD.ignoreEmptySplits` su `true` di default. Come soluzione alternativa, imposta esplicitamente `spark.hadoopRDD.ignoreEmptySplits` su `false`. Amazon EMR rilascio 6.9.0 risolve questo problema.
- Su cluster Trino a lunga esecuzione Amazon EMR 6.6.0 abilita i parametri di registrazione della rimozione di oggetti inutili (Garbage Collection) in Trino `jvm.config` per ottenere informazioni migliori dai log di Garbage Collection. Questa modifica aggiunge molti registri di Garbage Collection al file `launcher.log` (`/var/log/trino/launcher.log`). Se esegui cluster Trino in Amazon EMR 6.6.0, potresti riscontrare che i nodi esauriscono lo spazio su disco dopo che il cluster è stato in esecuzione per un paio di giorni a causa dei registri che sono stati aggiunti.

La soluzione alternativa per questo problema è eseguire lo script sottostante come operazione Bootstrap per disabilitare i parametri di registrazione di rimozione di oggetti inutili (garbage collection) in `jvm.config` durante la creazione o la clonazione del cluster per Amazon EMR 6.6.0.


```
#!/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"
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

- Con le release 5.36.0 e da 6.6.0 a 6.9.0 di Amazon EMR, i componenti dei servizi `SecretAgent` e `RecordServer` servizio potrebbero subire una perdita di dati di log a causa di una configurazione errata del modello nome file nelle proprietà `Log4j2`. Ca sua della configurazione errata, i componenti generano un solo file di log al giorno. Quando si applica la strategia di rotazione, sovrascrive il file esistente invece di generare un nuovo file di log come previsto. Come soluzione alternativa, utilizza un'azione bootstrap per generare file di log ogni ora e accoda un numero intero con incremento automatico nel nome file per gestire la rotazione.

Per le release da 6.6.0 a 6.9.0 di Amazon EMR, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Per Amazon EMR 5.36.0, utilizza la seguente azione bootstrap quando avvii un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Rilascio 5.35.0

Questa è la nota di rilascio di Amazon EMR 5.35.0.

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.35.0. Le modifiche sono relative alla versione 5.34.0.

Data del rilascio iniziale: 30 marzo 2022

Nuove caratteristiche

- Le applicazioni Amazon EMR versione 5.35 che usano Log4j 1.x e Log4j 2.x sono aggiornate per usare rispettivamente Log4j 1.2.17 (o superiore) e Log4j 2.17.1 (o superiore) e non richiedono l'uso di operazioni di bootstrap per mitigare i problemi CVE nelle versioni precedenti. Per informazioni, consultare [Approccio per mitigare il CVE-2021-44228](#).

Modifiche, miglioramenti e problemi risolti

Modifiche di Flink

Tipo di modifica	Descrizione
Aggiornamenti	<ul style="list-style-type: none"> Aggiorna la versione di flink alla 1.14.2. log4j aggiornato a 2.17.1.

Cambiamenti di Hadoop

Tipo di modifica	Descrizione
Backport open source di Hadoop da EMR 5.34.0	<ul style="list-style-type: none"> YARN-10438: Gestire null ContainerID in ClientRemService#getContainerReport() YARN-7266: thread del gestore eventi Timeline Server bloccati

Tipo di modifica	Descrizione
	<ul style="list-style-type: none"> • YARN-10438: ATS 1.5 non si avvia se i file RollingLevelDB sono danneggiati o mancanti • HADOOP-13500: Sincronizzazione dell'iterazione dell'oggetto delle proprietà di configurazione • YARN-10651: CapacityScheduler si è bloccato con NPE in AbstractYarnScheduler.updateNodeResource() • HDFS-12221: Sostituisci gli xerces in XmlEditsVisitor • HDFS-16410: analisi Xml non sicura in OfflineEditSXMLLoader
Modifiche e correzioni di Hadoop	<ul style="list-style-type: none"> • Tomcat utilizzato in KMS e HTTPFS è aggiornato a 8.5.75 • In FileSystemOptimizedCommitterV2, il marker di successo è stato scritto nel percorso di output CommitJob definito durante la creazione del committer. Poiché i percorsi di output CommitJob e livello di attività possono essere diversi, il percorso è stato corretto per utilizzare quello definito nei file manifest. Per i lavori Hive, ciò comporta che il marker di successo venga scritto correttamente durante l'esecuzione di operazioni come la partizione dinamica o UNION ALL.

Cambiamenti Hive

Tipo di modifica	Descrizione
Hive aggiornato a open source versione 2.3.9 , incluse queste correzioni JIRA	<ul style="list-style-type: none"> • HIVE-17155: findConfFile() in HiveConf.java presenta alcuni problemi con il percorso conf

Tipo di modifica	Descrizione
	<ul style="list-style-type: none"> • HIVE-24797: disabilita la convalida dei valori predefiniti durante l'analisi degli schemi Avro • HIVE-21563: Migliora le prestazioni di Table #getEmptyTable disabilita RegisterAllFunctionsOnce • HIVE-18147: i test possono fallire con java.net.bindeXception: indirizzo già in uso • HIVE-24608: torna a get_table nel client HMS per Hive 2.3.x • HIVE-21200: Vectorization - colonna data che lancia java.lang.UnsupportedOperationException per parquet • HIVE-19228: rimuovere l'utilizzo di commons-httpclient 3.x
Hive backport open source da EMR 5.34.0	<ul style="list-style-type: none"> • HIVE-19990: la query con intervallo letterale nella condizione di join non riesce • HIVE-25824: Aggiorna branch-2.3 a log4j 2.17.0 • TEZ-4062: la pianificazione dei tentativi speculativi deve essere interrotta al completamento dell'attività • TEZ-4108: NullPointerException durante la condizione di gara di esecuzione speculativa • TEZ-3918: L'impostazione di tez.task.log.level non funziona
Aggiornamenti e correzioni di Hive	<ul style="list-style-type: none"> • Aggiorna la versione di Log4j a 2.17.1 • Aggiorna la versione di ORC a 1.4.3 • Stallo fisso dovuto al thread di penalità in ShuffleScheduler

Tipo di modifica	Descrizione
Nuove funzionalità di	<ul style="list-style-type: none"> Aggiunta la funzione per stampare Hive Query nei registri AM. Questa opzione è disabilitata per impostazione predefinita. Bandiera/Conf:tez.am.emr.print.hive.query.in.log . Stato (predefinito): FALSE.

Modifiche Oozie

Tipo di modifica	Descrizione
Backport open source di Oozie da EMR 5.34.0	<ul style="list-style-type: none"> OOZIE-3652: Oozie launcher dovrebbe riprovare l'elenco delle directory quando si verifica NoSuchFileException

Cambiamenti Pig

Tipo di modifica	Descrizione
Aggiornamenti	<ul style="list-style-type: none"> log4j aggiornato a 1.2.17.

Problemi noti

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio s3://bucket/table/p=a è un prefisso di s3://bucket/table/p=a b.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere / (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio

(U+0020) che compare tra a e b in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 5.34.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.34.0. Le modifiche sono relative alla versione 5.33.1.

Data del rilascio iniziale: 20 gennaio 2022

Data di rilascio aggiornata: 21 marzo 2022

Nuove caratteristiche

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).
- [Hudi] Miglioramenti per semplificare la configurazione Hudi. Optimistic Concurrency Control disabilitato per impostazione predefinita.

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- In precedenza, il riavvio manuale del gestore delle risorse su un cluster multi-master provocava il ricaricamento, da parte dei daemon su cluster di Amazon EMR come Zookeeper, di tutti i nodi precedentemente disattivati o persi nel file `znode` di Zookeeper. Ciò causava il superamento dei limiti predefiniti in determinate situazioni. Ora Amazon EMR rimuove i record dei nodi disattivati o persi più vecchi di un'ora dal file Zookeeper e i limiti interni sono stati aumentati.

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Zeppelin aggiornato alla versione 0.10.0.
- Livy Fix: aggiornato alla versione 0.7.1
- Miglioramento delle prestazioni di Spark: gli esecutori eterogenei sono disabilitati quando alcuni valori di configurazione Spark vengono sovrascritti in EMR 5.34.0.
- I server HttpFS e WebHDFS sono disabilitati per impostazione predefinita. Puoi abilitare nuovamente WebHDFS utilizzando la configurazione di Hadoop, `dfs.webhdfs.enabled`. Il server HttpFS può essere avviato utilizzando `sudo systemctl start hadoop-httpfs`.

Problemi noti

- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.

- Le query Hue non funzionano in Amazon EMR 6.4.0 perché il server HttpFS di Apache Hadoop è disabilitato per impostazione predefinita. Per utilizzare Hue su Amazon EMR 6.4.0, avvia manualmente il server HttpFS sul nodo primario di Amazon EMR utilizzando `sudo systemctl start hadoop-httpfs` oppure [utilizza una fase di Amazon EMR](#).
- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 6.5.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.5.0. Le modifiche sono relative alla versione 6.4.0.

Data del rilascio iniziale: 20 gennaio 2022

Data di rilascio aggiornata: 21 marzo 2022

Nuove caratteristiche

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).
- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.
- Supporto per il formato di tabella aperta Apache Iceberg per enormi set di dati analitici.
- Supporto per ranger-trino-plugin 2.0.1-amzn-1
- Supporto per toree 0.5.0

Modifiche, miglioramenti e problemi risolti

- La versione del rilascio 6.5 di Amazon EMR ora supporta Apache Iceberg 0.12.0 e offre miglioramenti al tempo di esecuzione con Amazon EMR Runtime per Apache Spark, Amazon EMR Runtime per Presto e Amazon EMR Runtime per Apache Hive.
- [Apache Iceberg](#) è un formato di tabella aperta per set di dati di grandi dimensioni in Amazon S3 e fornisce prestazioni di query rapide su tabelle di grandi dimensioni, commit atomici, scritture simultanee ed evoluzione delle tabelle compatibili con SQL. Con EMR 6.5, è possibile utilizzare Apache Spark 3.1.2 con il formato tabella Iceberg.
- Apache Hudi 0.9 aggiunge il supporto Spark SQL DDL e DML. Ciò consente di creare e modificare le tabelle Hudi usando solo istruzioni SQL. Apache Hudi 0.9 include anche miglioramenti delle prestazioni sul lato query e sul lato scrittore.
- Amazon EMR Runtime per Apache Hive migliora le prestazioni di Apache Hive su Amazon S3 rimuovendo le operazioni di ridenominazione durante le operazioni di gestione temporanea e migliora le prestazioni per i comandi di metastore check (MSCK) utilizzati per la riparazione delle tabelle.

Problemi noti

- Quando le release 6.5.0, 6.6.0 o 6.7.0 di Amazon EMR leggono le tabelle di Apache Phoenix tramite la shell (interprete di comandi) Apache Spark, si verifica un `NoSuchMethodError` perché Amazon EMR utilizza un valore errato `Hbase.compat.version`. Il rilascio 6.8.0 di Amazon EMR risolve questo problema.
- I cluster bundle Hbase in High Availability (HA) non riescono a effettuare il provisioning con la dimensione del volume e il tipo di istanza di default. La soluzione alternativa per questo problema consiste nell'aumentare la dimensione del volume principale.
- Per utilizzare le operazioni Spark con Apache Oozie, devi aggiungere la seguente configurazione al tuo file `workflow.xml` Oozie. In caso contrario, diverse librerie critiche come Hadoop e EMRFS non saranno presenti nella classpath degli esecutori Spark lanciati da Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 6.4.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.4.0. Le modifiche sono relative alla versione 6.3.0.

Data del rilascio iniziale: 20 settembre 2021

Data di rilascio aggiornata: 21 marzo 2022

Applicazioni supportate

- AWS SDK for Java versione 1.12.31
- CloudWatch Sink versione 2.2.0
- DynamoDB Connector versione 4.16.0
- EMRFS versione 2.47.0
- Amazon EMR Goodies versione 3.2.0
- Amazon EMR Kinesis Connector versione 3.5.0
- Amazon EMR Record Server versione 2.1.0
- Amazon EMR Scripts versione 2.5.0
- Flink versione 1.13.1
- Ganglia versione 3.7.2
- AWS Glue Hive Metastore Client versione 3.3.0
- Hadoop versione 3.2.1-amzn-4
- HBase versione 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalog versione 3.1.2-amzn-5
- Hive versione 3.1.2-amzn-5
- Hudi versione 0.8.0-amzn-0
- Hue versione 4.9.0
- Java JDK versione Corretto-8.302.08.1 (build 1.8.0_302-b08)
- JupyterHub versione 1.4.1
- Livy versione 0.7.1-incubating
- MXNet versione 1.8.0
- Oozie versione 5.2.1

- Phoenix versione 5.1.2
- Pig versione 0.17.0
- Presto versione 0.254.1-amzn-0
- Trino versione 359
- Apache Ranger KMS (crittografia trasparente multi-master) versione 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK versione 1.4.1
- Scala versione 2.12.10 (VM server OpenJDK a 64 bit, Java 1.8.0_282)
- Spark versione 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop versione 1.4.7
- TensorFlow versione 2.4.1
- tez versione 0.9.2
- Zeppelin versione 0.9.0
- Zookeeper versione 3.5.7
- Connettori e driver: DynamoDB Connector 4.16.0

Nuove funzionalità di

- [Dimensionamento] Ottimizzazione del dimensionamento gestito dei dati di shuffle di Spark - Per Amazon EMR versioni 5.34.0 e successive e EMR 6.4.0 e successive, il dimensionamento gestito è ora consapevole dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta [Utilizzo del dimensionamento gestito da EMR in Amazon EMR](#) nella Guida di gestione di Amazon EMR e [Guida di programmazione Spark](#).
- Sui cluster Amazon EMR abilitati ad Apache Ranger, puoi utilizzare Apache Spark SQL per inserire o aggiornare i dati nelle tabelle dei metastore di Apache Hive utilizzando INSERT INTO, INSERT OVERWRITE, e ALTER TABLE. Quando si utilizza ALTER TABLE con Spark SQL, una posizione di partizione deve essere la directory figlio di una posizione di tabella. Al momento Amazon EMR non supporta l'inserimento di dati in una partizione in cui la posizione della partizione è diversa da quella della tabella.
- PrestoSQL è stato [rinominato in Trino](#).

- Hive: l'esecuzione di semplici query SELECT con clausola LIMIT viene accelerata interrompendo l'esecuzione della query non appena viene recuperato il numero di registri menzionati nella clausola LIMIT. Le query SELECT semplici sono query che non hanno una clausola GROUP BY/ORDER by o query che non hanno uno stadio di riduzione. Ad esempio, `SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>`.

Controlli di simultaneità Hudi

- Hudi ora supporta Optimistic Concurrency Control (OCC), che può essere sfruttato con operazioni di scrittura come UPSERT e INSERT per consentire modifiche da più scrittori alla stessa tabella Hudi. Questo è livello di file OCC, quindi due commit (o scrittori) possono scrivere sulla stessa tabella, se le modifiche non sono in conflitto. Per ulteriori informazioni, consulta [Controllo della concorrenza Hudi](#).
- I cluster Amazon EMR hanno installato Zookeeper, che può essere sfruttato come provider di blocchi per OCC. Per semplificare l'utilizzo di questa funzione, i cluster Amazon EMR hanno le seguenti proprietà preconfigurate:

```
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
```

Per abilitare OCC, è necessario configurare le seguenti proprietà con le relative opzioni di processo Hudi o a livello di cluster utilizzando l'API di configurazione Amazon EMR:

```
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)
```

Monitoraggio Hudi: integrazione di Amazon CloudWatch per segnalare i parametri Hudi

- Amazon EMR supporta la pubblicazione di parametri Hudi su Amazon CloudWatch. È abilitato impostando le seguenti configurazioni richieste:

```
hoodie.metrics.on=true
```

```
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Di seguito sono riportate le configurazioni Hudi opzionali che è possibile modificare:

Impostazione	Descrizione	Valore
hoodie.metrics.cloudwatch.report.period.seconds	Frequenza (in secondi) con cui segnalare i parametri ad Amazon CloudWatch	Il valore di default è 60s, il che va bene per la risoluzione di default di un minuto offerta da Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefisso da aggiungere a ciascun nome parametro	Il valore di default è vuoto (nessun prefisso)
hoodie.metrics.cloudwatch.namespace	Spazio dei nomi Amazon CloudWatch in base al quale vengono pubblicati i parametri	Il valore di default è Hudi
hoodie.metrics.cloudwatch.maxDatumsPerRequest	Numero massimo di riferimenti da includere in una richiesta ad Amazon CloudWatch	Il valore di default è 20, uguale a quello di default di Amazon CloudWatch

Supporto e miglioramenti delle configurazioni Amazon EMR Hudi

- I clienti possono ora sfruttare l'API di configurazione EMR e la funzione di riconfigurazione per configurare le configurazioni Hudi a livello di cluster. Un nuovo supporto per la configurazione basato su file è stato introdotto tramite `/etc/hudi/conf/hudi-defaults.conf` sulla linea di altre applicazioni come Spark, Hive ecc. EMR configura alcuni valori di default per migliorare l'esperienza utente:

— `hoodie.datasource.hive_sync.jdbcurl` è configurato per l'URL del server Hive del cluster e non deve più essere specificato. Ciò è particolarmente utile quando si esegue un processo in modalità cluster Spark, dove in precedenza era necessario specificare l'IP principale Amazon EMR.

— Configurazioni specifiche di HBase, utili per l'utilizzo dell'indice HBase con Hudi.

- Configurazione specifica del provider di blocco Zookeeper, come discusso sotto il controllo della concorrenza, che semplifica l'utilizzo di Optimistic Concurrency Control (OCC).
- Sono state introdotte ulteriori modifiche per ridurre il numero di configurazioni che devi passare e per dedurre automaticamente laddove possibile:
 - La parola chiave `partitionBy` può essere utilizzata per specificare la colonna della partizione.
 - Quando si abilita Hive Sync, non è più obbligatorio passare `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Tali valori possono essere dedotti dal nome della tabella Hudi e dal campo della partizione.
 - `KEYGENERATOR_CLASS_OPT_KEY` non è obbligatorio passare e può essere dedotto da casi più semplici di `SimpleKeyGenerator` e `ComplexKeyGenerator`.

Avvertenze Hudi

- Hudi non supporta l'esecuzione vettorizzata nelle tabelle Hive for Merge on Read (MoR) e Bootstrap. Ad esempio: `count(*)` fallisce con la tabella in tempo reale Hudi quando `hive.vectorized.execution.enabled` è impostato su `VERO`. Come soluzione alternativa, puoi disabilitare la lettura vettorizzata impostando `hive.vectorized.execution.enabled` a `false`.
- Il supporto multi-writer non è compatibile con la caratteristica bootstrap Hudi.
- Flink Streamer e Flink SQL sono caratteristiche sperimentali di questa versione. Queste caratteristiche non sono consigliate per l'uso nelle distribuzioni di produzione.

Modifiche, miglioramenti e problemi risolti

Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.

- In precedenza, il riavvio manuale del gestore delle risorse su un cluster multi-master provocava il ricaricamento, da parte dei daemon su cluster di Amazon EMR come Zookeeper, di tutti i nodi precedentemente disattivati o persi nel file `znode` di Zookeeper. Ciò causava il superamento dei limiti predefiniti in determinate situazioni. Ora Amazon EMR rimuove i record dei nodi disattivati o persi più vecchi di un'ora dal file Zookeeper e i limiti interni sono stati aumentati.

- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Configurazione di un cluster per risolvere i problemi di prestazioni di Apache YARN Timeline Server versione 1 e 1.5

Le versioni 1 e 1.5 di Apache YARN Timeline Server possono causare problemi di prestazioni con cluster EMR molto attivi e di grandi dimensioni, in particolare con `yarn.resourcemanager.system-metrics-publisher.enabled=true`, che è l'impostazione predefinita in Amazon EMR. Un YARN Timeline Server v2 open source risolve il problema di prestazioni relativo alla scalabilità di YARN Timeline Server.

Altre soluzioni alternative per questo problema includono:

- La configurazione di `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- L'abilitazione della correzione per questo problema durante la creazione di un cluster, come descritto di seguito.

I seguenti rilasci di Amazon EMR contengono una correzione per questo problema di prestazioni di YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1 e 6.4.x

Per abilitare la correzione su una delle versioni di Amazon EMR sopra specificate, imposta queste proprietà su `true` in un file JSON di configurazioni che viene passato utilizzando il [parametro del comando `aws emr create-cluster`](#): `--configurations file:///./configurations.json`. Oppure abilita la correzione utilizzando l'[interfaccia utente della console di riconfigurazione](#).

Esempio del contenuto del file `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": []
}
]
```

- I server HttpFS e WebHDFS sono disabilitati per impostazione predefinita. Puoi abilitare nuovamente WebHDFS utilizzando la configurazione di Hadoop, `dfs.webhdfs.enabled`. Il server HttpFS può essere avviato utilizzando `sudo systemctl start hadoop-httpfs`.
- Il protocollo HTTPS ora è abilitato per impostazione predefinita per i repository Amazon Linux. Se utilizzi una policy Amazon S3 VPCE per limitare l'accesso a specifici bucket, devi aggiungere il nuovo ARN del bucket Amazon Linux `arn:aws:s3:::amazonlinux-2-repos-$region/*` alla tua policy (sostituisci `$region` con la Regione in cui si trova l'endpoint). Per ulteriori informazioni, consulta questo argomento nella forum di discussione di AWS. [Annuncio: Amazon Linux 2 ora supporta la possibilità di utilizzare HTTPS durante la connessione ai repository dei pacchetti](#).
- Hive: le prestazioni delle query di scrittura sono migliorate abilitando l'uso di una directory Scratch su HDFS per l'ultimo processo. I dati temporanei per il processo finale vengono scritti su HDFS anziché su Amazon S3 e le prestazioni sono migliorate perché i dati vengono spostati da HDFS alla posizione finale della tabella (Amazon S3) anziché tra i dispositivi Amazon S3.
- Hive: miglioramento del tempo di compilazione delle query fino a 2,5 volte con Glue metastore Partition Pruning.

- Per impostazione predefinita, quando le FDU integrate vengono passate da Hive al Hive Metastore Server, solo un sottoinsieme di tali FDU incorporati viene passato al Glue Metastore poiché Glue supporta solo operatori di espressioni limitate. Se hai impostato `hive.glue.partition.pruning.client=true`, quindi tutta la eliminazione delle partizioni avviene sul lato client. Se hai impostato `hive.glue.partition.pruning.server=true`, quindi tutta la eliminazione delle partizioni avviene sul lato server.

Problemi noti

- Le query Hue non funzionano in Amazon EMR 6.4.0 perché il server HttpFS di Apache Hadoop è disabilitato per impostazione predefinita. Per utilizzare Hue su Amazon EMR 6.4.0, avvia manualmente il server HttpFS sul nodo primario di Amazon EMR utilizzando `sudo systemctl start hadoop-httpfs` oppure [utilizza una fase di Amazon EMR](#).
- La caratteristica di Amazon EMR Notebooks utilizzata con la rappresentazione utente di Livy non funziona perché HttpFS è disabilitato per impostazione predefinita. In questo caso, EMR Notebooks non può connettersi al cluster in cui è abilitata la rappresentazione di Livy. La soluzione alternativa consiste nell'avviare il server HttpFS prima di connettere EMR Notebooks al cluster utilizzando `sudo systemctl start hadoop-httpfs`.
- In Amazon EMR versione 6.4.0, Phoenix non supporta il componente dei connettori Phoenix.
- Per utilizzare le operazioni Spark con Apache Oozie, devi aggiungere la seguente configurazione al tuo file `workflow.xml` Oozie. In caso contrario, diverse librerie critiche come Hadoop e EMRFS non saranno presenti nella classpath degli esecutori Spark lanciati da Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono

altri 14 caratteri non di controllo: !"#%&'()*+,-. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 5.32.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.32.0. Le modifiche sono relative alla versione 5.31.0.

Data del rilascio iniziale: 8 gennaio 2021

Aggiornamenti

- Connettore Amazon Glue aggiornato alla versione 1.14.0
- Aggiornamento di Amazon SageMaker Spark SDK alla versione 1.4.1
- Aggiornato AWS SDK for Java alla versione 1.11.890
- EMR DynamoDB Connector aggiornato alla versione 4.16.0
- EMRFS aggiornato alla versione 2.45.0
- Parametri di analisi dei log EMR aggiornati alla versione 1.18.0
- Client MetricsAndEventsApiGateway EMR aggiornato alla versione 1.5.0
- EMR Record Server aggiornato alla versione 1.8.0
- EMR S3 Dist CP aggiornato alla versione 2.17.0
- EMR Secret Agent aggiornato alla versione 1.7.0
- Flink aggiornato alla versione 1.11.2
- Hadoop aggiornato alla versione 2.10.1-amzn-0
- Hive aggiornato alla versione 2.3.7-amzn-3
- Hue aggiornato alla versione 4.8.0
- MXNet aggiornato alla versione 1.7.0
- OpenCV aggiornato alla versione 4.4.0
- Presto aggiornato alla versione 0.240.1-amzn-0
- Spark aggiornato alla versione 2.4.7-amzn-0

- TensorFlow aggiornato alla versione 2.3.1

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Versioni dei componenti aggiornate.
- Per un elenco delle versioni dei componenti, consulta [Informazioni sui rilasci di Amazon EMR](#) in questa guida.

Nuove funzionalità di

- A partire da Amazon EMR 5.32.0 e 6.5.0, il ridimensionamento dell'executor dinamico per Apache Spark è abilitato di default. Per attivare o disattivare questa caratteristica, è possibile utilizzare il `spark.yarn.heterogeneousExecutors.enabled` parametro di configurazione.
- Stato del supporto IMDS (Instance Metadata Service) V2: Amazon EMR 5.23.1, 5.27.1 e 5.32 o versioni successive utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Per altri rilasci EMR 5.x, la disattivazione di IMDSv1 causa un errore di avvio del cluster.
- A partire da Amazon EMR 5.32.0, è possibile avviare un cluster che si integra nativamente con Apache Ranger. Apache Ranger è un framework open source che consente di abilitare, monitorare e gestire la sicurezza completa dei dati attraverso la piattaforma Hadoop. Per ulteriori informazioni, consulta [Apache Ranger](#). L'integrazione nativa consente di utilizzare Apache Ranger per imporre un controllo granulare di accesso ai dati su Amazon EMR. Consulta [Integrazione di Amazon EMR con Apache Ranger](#) nella Guida ai rilasci di Amazon EMR.
- Amazon EMR 5.32.0 supporta Amazon EMR su EKS. Per ulteriori dettagli su come iniziare a utilizzare EMR su EKS, consulta [Che cos'è Amazon EMR su EKS?](#).
- Amazon EMR 5.32.0 supporta Amazon EMR Studio (anteprima). Per ulteriori informazioni sulle nozioni di base di EMR Studio, consulta [Amazon EMR Studio \(anteprima\)](#).
- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).

Problemi noti

- Per i cluster della sottorete privata di Amazon EMR 6.3.0 e 6.2.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio httpd sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.
- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno

un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
```

```
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Il supporto della console per creare una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger non è attualmente abilitato nella Regione GovCloud. La configurazione della sicurezza può essere eseguita utilizzando la CLI. Consulta [Creazione della configurazione di sicurezza EMR](#) nella Guida per la gestione di Amazon EMR.
- Quando la crittografia AtRestEncryption o HDFS è abilitata in un cluster che utilizza Amazon EMR 5.31.0 o 5.32.0, le query Hive determinano la seguente eccezione di runtime.

```
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
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 6.2.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.2.0. Le modifiche sono relative alla versione 6.1.0.

Data del rilascio iniziale: 9 dicembre 2020

Ultimo aggiornamento: 4 ottobre 2021

Applicazioni supportate

- AWS SDK for Java versione 1.11.828
- emr-record-server versione 1.7.0
- Flink versione 1.11.2
- Ganglia versione 3.7.2
- Hadoop versione 3.2.1-amzn-1

- HBase versione 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog versione 3.1.2-amzn-0
- Hive versione 3.1.2-amzn-3
- Hudi versione 0.6.0-amzn-1
- Hue versione 4.8.0
- JupyterHub versione 1.1.0
- Livy versione 0.7.0
- MXNet versione 1.7.0
- Oozie versione 5.2.0
- Phoenix versione 5.0.0
- Pig versione 0.17.0
- Presto versione 0.238.3-amzn-1
- PrestoSQL versione 343
- Spark versione 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow versione 2.3.1
- Zeppelin versione 0.9.0-preview1
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.16.0

Nuove funzionalità di

- HBase: rimossa la rinomina in fase di commit e aggiunto il tracciamento HFile persistente. Consulta [Tracciamento HFile persistente](#) nella Guida ai rilasci di Amazon EMR.
- HBase: backport per [Creazione di una configurazione che costringe a memorizzare nella cache i blocchi sulla compattazione](#).
- PrestoDB: miglioramenti all'eliminazione delle partizioni dinamiche. Join Reorder basato su regole funziona su dati non partizionati.
- Criteri gestiti con ambito: per allinearsi con le best practice di AWS, Amazon EMR ha introdotto le policy con ambito gestite di default EMR v2 come sostituzioni per le policy che saranno rese obsolete. Consulta [Policy gestite di Amazon EMR](#).

- Stato del supporto IMDS (Instance Metadata Service) V2: per Amazon EMR 6.2 o versioni successive, i componenti Amazon EMR utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Se si disabilita IMDSv1 nelle versioni precedenti di Amazon EMR 6.x, si verificherà un errore di avvio del cluster.

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Spark: miglioramenti delle prestazioni del runtime Spark.

Problemi noti

- Amazon EMR 6.2 dispone di autorizzazioni errate impostate sul file `/etc/cron.d/libinstance-controller-java` in EMR 6.2.0. Le autorizzazioni sul file sono `645 (-rw-r--r-x)`, quando dovrebbero essere `644 (-rw-r--r--)`. Di conseguenza, Amazon EMR versione 6.2 non registra i log dello stato delle istanze e la directory `/emr/instance-logs` è vuota. Questo problema è stato risolto in Amazon EMR 6.3.0 e successive.

Per risolvere il problema, esegui il seguente script come operazione di bootstrap all'avvio del cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Per i cluster della sottorete privata di Amazon EMR 6.2.0 e 6.3.0 non è possibile accedere all'interfaccia utente Web di Ganglia. Verrà visualizzato l'errore "access denied (403) (accesso negato (403))". Altre interfacce utente Web, come Spark, Hue, JupyterHub, Zeppelin, Livy e Tez funzionano normalmente. Anche l'accesso all'interfaccia utente Web di Ganglia sui cluster della sottorete pubblica funziona normalmente. Per risolvere il problema, riavvia il servizio `httpd` sul nodo primario con `sudo systemctl restart httpd`. Questo problema è stato risolto in Amazon EMR 6.4.0.
- C'è un problema in Amazon EMR 6.2.0 in cui `httpd` ha continuamente esito negativo, causando la non disponibilità di Ganglia. Viene visualizzato un errore "cannot connect to the server (impossibile connettersi al server)". Per risolvere questo problema in un cluster che è già in esecuzione, abilita SSH sul nodo primario del cluster e aggiungi la riga `Listen 80` al file `httpd.conf` presente in `/etc/httpd/conf/httpd.conf`. Questo problema è stato risolto in Amazon EMR 6.3.0.
- HTTPD ha esito negativo su cluster EMR 6.2.0 quando si utilizza una configurazione di sicurezza. Ciò rende l'interfaccia utente dell'applicazione Web Ganglia non disponibile. Per accedere all'interfaccia utente dell'applicazione Web Ganglia, aggiungi `Listen 80` al file `/etc/httpd/conf/httpd.conf` sul nodo primario del cluster. Per ulteriori informazioni su come connettersi al cluster, consulta la sezione [Connect to the Primary Node Using SSH](#) (Connessione al nodo primario tramite SSH).

Inoltre, EMR Notebooks non riesce a stabilire una connessione con cluster EMR 6.2.0 quando si utilizza una configurazione di sicurezza. Il notebook non riuscirà a elencare i kernel e a inviare processi Spark. Ti consigliamo invece di utilizzare EMR Notebooks con un'altra versione di Amazon EMR.

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
```

```
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **⚠ Important**

Amazon EMR 6.1.0 e 6.2.0 includono un problema di prestazioni che può influire in modo critico su tutte le operazioni di inserimento, upsert ed eliminazione di Hudi. Se intendi utilizzare Hudi con Amazon EMR 6.1.0 o 6.2.0, contatta il supporto AWS per ottenere un RPM Hudi con patch.

- **⚠ Important**

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi [personalizzare l'AMI di Amazon EMR](#) per [utilizzare l'AMI di Amazon Linux più recente](#).

- Gli artifact di Amazon EMR 6.2.0 Maven non vengono pubblicati. Saranno pubblicati con una futura versione di Amazon EMR.

- Il tracciamento HFile persistente che utilizza la tabella di sistema Storefile di HBase non supporta la funzionalità di replica della regione HBase. Per ulteriori informazioni sulla replica della regione HBase, consulta [Lecture ad alta disponibilità coerenti con la tempistica](#).
- Differenze di versione dei bucket Hive tra Amazon EMR 6.x e EMR 5.x

EMR 5.x utilizza OOS Apache Hive 2, mentre EMR 6.x utilizza OOS Apache Hive 3. L'Hive2 open source utilizza bucket di versione 1, mentre l'Hive3 open source utilizza bucket di versione 2. Questa differenza di versione dei bucket tra Hive 2 (EMR 5.x) e Hive 3 (EMR 6.x) presuppone un funzionamento diverso dell'hashing dei bucket Hive. Vedere l'esempio sottostante.

La tabella seguente è un esempio creato rispettivamente in EMR 6.x ed EMR 5.x.

```
-- 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/';
```

Sono stati inseriti gli stessi dati in EMR 6.x ed EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Osservando la posizione S3, si nota che il nome del file di bucket è diverso, poiché la funzione di hashing è diversa tra EMR 6.x (Hive 3) e EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

Puoi notare la differenza di versione anche eseguendo il comando riportato di seguito nella CLI Hive in EMR 6.x. Potrai notare che la CLI restituisce bucket di versione 2.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.

- Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 5.31.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.31.0. Le modifiche sono relative alla versione 5.30.1.

Data del rilascio iniziale: 9 ottobre 2020

Ultimo aggiornamento: 15 ottobre 2020

Aggiornamenti

- Connettore Amazon Glue aggiornato alla versione 1.13.0
- Amazon SageMaker Spark SDK aggiornato alla versione 1.4.0
- Connettore Amazon Kinesis aggiornato alla versione 3.5.9
- Aggiornato AWS SDK for Java alla versione 1.11.852
- Bigtop-tomcat aggiornato alla versione 8.5.56
- EMR FS aggiornato alla versione 2.43.0
- Client MetricsAndEventsApiGateway EMR aggiornato alla versione 1.4.0
- EMR S3 Dist CP aggiornato alla versione 2.15.0
- EMR S3 Select aggiornato alla versione 1.6.0
- Flink aggiornato alla versione 1.11.0
- Hadoop aggiornato alla versione 2.10.0

- Hive aggiornato alla versione 2.3.7
- Hudi aggiornato alla versione 0.6.0
- Hue aggiornato alla versione 4.7.1
- Aggiornato JupyterHub alla versione 1.1.0
- MXNet aggiornato alla versione 1.6.0
- OpenCV aggiornato alla versione 4.3.0
- Presto aggiornato alla versione 0.238.3
- TensorFlow aggiornato alla versione 2.1.0

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1,

5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).

- Le [statistiche di colonna Hive](#) sono supportate per Amazon EMR 5.31.0 e versioni successive.
- Versioni dei componenti aggiornate.
- Supporto di EMRFS S3EC V2 in Amazon EMR 5.31.0. In S3 Java SDK 1.11.837 e versioni successive, il client di crittografia versione 2 (S3EC V2) è stato introdotto con vari miglioramenti alla sicurezza. Per ulteriori informazioni, consulta gli argomenti seguenti:
 - Post del blog S3: [Updates to the Amazon S3 encryption client \(Aggiornamenti al client di crittografia Amazon S3\)](#).
 - AWS SDK for Java Guida per gli sviluppatori: [Migrazione dei client di crittografia e decrittografia a V2](#).
 - Guida alla gestione di EMR: [Crittografia lato client di Amazon S3](#).

Il client di crittografia V1 è ancora disponibile nell'SDK per la compatibilità con le versioni precedenti.

Nuove funzionalità di

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Con Amazon EMR 5.31.0, puoi avviare un cluster che si integra con Lake Formation. L'integrazione fornisce un filtraggio dei dati granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Inoltre, consente l'accesso federato single sign-on a EMR Notebooks o Apache Zeppelin

da un sistema di identità aziendale. Per ulteriori informazioni, consulta [Integrazione di Amazon EMR con AWS Lake Formation](#) nella Guida alla gestione di Amazon EMR.

Amazon EMR con Lake Formation è attualmente disponibile in 16 regioni AWS: Stati Uniti orientali (Ohio e N. Virginia), Stati Uniti occidentali (California settentrionale e Oregon), Asia Pacifico (Mumbai, Seoul, Singapore, Sydney e Tokyo), Canada (Centrale), Europa (Francoforte, Irlanda, Londra, Parigi e Stoccolma), Sud America (San Paolo).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Quando la crittografia AtRestEncryption o HDFS è abilitata in un cluster che utilizza Amazon EMR 5.31.0 o 5.32.0, le query Hive determinano la seguente eccezione di runtime.

```
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
```

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-`. Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 6.1.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.1.0. Le modifiche sono relative alla versione 6.0.0.

Data del rilascio iniziale: 4 settembre 2020

Ultimo aggiornamento: 15 ottobre 2020

Applicazioni supportate

- AWS SDK for Java versione 1.11.828
- Flink versione 1.11.0

- Ganglia versione 3.7.2
- Hadoop versione 3.2.1-amzn-1
- HBase versione 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog versione 3.1.2-amzn-0
- Hive versione 3.1.2-amzn-1
- Hudi versione 0.5.2 in fase di sviluppo
- Hue versione 4.7.1
- JupyterHub versione 1.1.0
- Livy versione 0.7.0
- MXNet versione 1.6.0
- Oozie versione 5.2.0
- Phoenix versione 5.0.0
- Presto versione 0.232
- PrestoSQL versione 338
- Spark versione 3.0.0-amzn-0
- TensorFlow versione 2.1.0
- Zeppelin versione 0.9.0-preview1
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.14.0

Nuove funzionalità di

- I tipi di istanze ARM sono supportati a partire da Amazon EMR versione 5.30.0 e Amazon EMR versione 6.1.0.
- I tipi di istanze per uso generico M6g sono supportati a partire dalle versioni di Amazon EMR 6.1.0 e 5.30.0. Per ulteriori informazioni, consulta [Tipi di istanza supportati](#) nella Guida alla gestione di Amazon EMR.
- La caratteristica del gruppo di collocamento EC2 è supportata a partire da Amazon EMR versione 5.23.0 come opzione per i cluster con più nodi primari. Attualmente, solo i tipi di nodi primari sono supportati dalla caratteristica del gruppo di collocamento e la strategia SPREAD viene applicata a tali nodi primari. La strategia SPREAD colloca un piccolo gruppo di istanze su hardware

sottostante separato per evitare la perdita di più nodi primari in caso di guasto hardware. Per ulteriori informazioni, consulta [Integrazione di EMR con il gruppo di collocamento EC2](#) nella Guida alla gestione di Amazon EMR.

- Scalabilità gestita: con la versione 6.1.0 di Amazon EMR, puoi abilitare Scalabilità gestita di Amazon EMR per aumentare o diminuire automaticamente il numero di istanze o unità nel cluster in base al carico di lavoro. Amazon EMR valuta continuamente i parametri dei cluster per prendere decisioni di dimensionamento che ottimizzano i cluster in termini di costi e velocità. Il dimensionamento gestito è disponibile anche su Amazon EMR versione 5.30.0 e successive, tranne 6.0.0. Per maggiori informazioni, consulta [Dimensionamento delle risorse del cluster](#) nella Guida alla gestione di Amazon EMR.
- PrestoSQL versione 338 è supportata con EMR 6.1.0. Per ulteriori informazioni, consulta [Presto](#).
 - PrestoSQL è supportata solo su EMR 6.1.0 e versioni successive, non su EMR 6.0.0 o EMR 5.x.
 - Il nome dell'applicazione Presto continua ad essere utilizzato per installare PrestoDB sui cluster. Per installare PrestoSQL sui cluster, utilizza il nome dell'applicazione PrestoSQL.
 - È possibile installare PrestoDB o PrestoSQL, ma non è possibile installare entrambi in un singolo cluster. Se durante il tentativo di creare un cluster vengono specificati sia PrestoDB che PrestoSQL, si verifica un errore di convalida e la richiesta di creazione del cluster ha esito negativo.
 - PrestoSQL è supportato su entrambi i cluster single-master e multi-master. Nei cluster multi-master, è necessario un metastore Hive esterno per eseguire PrestoSQL o PrestoDB. Consulta la sezione [Applicazioni supportate in un cluster EMR con più nodi primari](#).
- Supporto per l'autenticazione automatica ECR su Apache Hadoop e Apache Spark con Docker: gli utenti di Docker possono utilizzare le immagini Docker da Docker Hub e Amazon Elastic Container Registry (Amazon ECR) per definire le dipendenze dell'ambiente e della libreria.

Consulta [Configurazione di Docker](#) ed [Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.x](#).

- EMR supporta le transazioni Apache Hive ACID: Amazon EMR 6.1.0 aggiunge il supporto per le transazioni Hive ACID in modo che sia conforme alle proprietà ACID di un database. Con questa funzione, puoi eseguire operazioni INSERT, UPDATE, DELETE, e MERGE nelle tabelle gestite da Hive con dati in Amazon Simple Storage Service (Amazon S3). Questa è una caratteristica chiave per casi d'uso come l'importazione dati in streaming, il riavvio dei dati, gli aggiornamenti in blocco tramite MERGE e la graduale variazione delle dimensioni. Per ulteriori informazioni, inclusi esempi di configurazione e casi d'uso, consulta [Amazon EMR supporta le transazioni Apache Hive ACID](#).

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.
- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Apache Flink non è supportata su EMR 6.0.0, ma è supportata su EMR 6.1.0 con Flink 1.11.0. Si tratta della prima versione di Flink a supportare ufficialmente Hadoop 3. Consulta [Annuncio del rilascio di Apache Flink 1.11.0](#).
- Ganglia è stato rimosso dai pacchetti EMR 6.1.0 di default.

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 e 6.2.0 includono un problema di prestazioni che può influire in modo critico su tutte le operazioni di inserimento, upsert ed eliminazione di Hudi. Se intendi utilizzare Hudi con Amazon EMR 6.1.0 o 6.2.0, contatta il supporto AWS per ottenere un RPM Hudi con patch.

- L'impostazione di configurazioni di garbage collection personalizzate con `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions` genera un errore di avvio del driver o dell'executor con EMR 6.1 a causa di una configurazione di garbage collection in conflitto. Con il rilascio EMR 6.1.0, è invece necessario specificare la configurazione di garbage collection Spark personalizzata per driver ed executor con le proprietà `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Per maggiori informazioni, consulta [Ambiente di runtime Apache Spark](#) e [Configurazione di garbage collection Spark su Amazon EMR 6.1.0](#).
- L'utilizzo di Pig con Oozie (e all'interno di Hue, dal momento che Hue utilizza le operazioni Oozie per eseguire gli script Pig) genera un errore che impedisce il caricamento di una libreria native-izo. Questo messaggio di errore è informativo e non impedisce l'esecuzione di Pig.
- Supporto della simultaneità di Hudi: attualmente, le scritture simultanee in una singola tabella Hudi non sono supportate. Inoltre, Hudi esegue il rollback di tutte le modifiche apportate dalle

istanze di scrittura in corso prima di consentire l'avvio di una nuova istanza di scrittura. Le scritture simultanee possono interferire con questo meccanismo e introdurre condizioni di competizione, le quali possono causare il danneggiamento dei dati. È necessario assicurarsi che, come parte del flusso di lavoro di elaborazione dei dati, ci sia sempre una sola istanza di scrittura Hudi che opera su una tabella Hudi. Hudi supporta più istanze di lettura simultanee che operano sulla stessa tabella Hudi.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Si è verificato un problema in Amazon EMR 6.1.0 che interessa i cluster che eseguono Presto. Dopo un lungo periodo di tempo (giorni), il cluster potrebbe generare errori come "su: failed to execute /bin/bash: Resource temporarily unavailable (su: impossibile eseguire /bin/bash: risorsa temporaneamente non disponibile)" o "shell request failed on channel 0 (richiesta shell non riuscita sul canale 0)". Questo problema è causato da un processo interno di Amazon EMR

(InstanceController) che genera un numero eccessivo di processi leggeri (LWP) e, alla fine, fa sì che l'utente Hadoop superi il limite nproc. Ciò impedisce all'utente di aprire ulteriori processi. La soluzione a questo problema consiste nell'eseguire l'aggiornamento a EMR 6.2.0.

Rilascio 6.0.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 6.0.0.

Data del rilascio iniziale: 10 marzo 2020

Applicazioni supportate

- AWS SDK for Java versione 1.11.711
- Ganglia versione 3.7.2
- Hadoop versione 3.2.1
- HBase versione 2.2.3
- HCatalog versione 3.1.2
- Hive versione 3.1.2
- Hudi versione 0.5.0 in fase di sviluppo
- Hue versione 4.4.0
- JupyterHub versione 1.0.0
- Livy versione 0.6.0
- MXNet versione 1.5.1
- Oozie versione 5.1.0
- Phoenix versione 5.0.0
- Presto versione 0.230
- Spark versione 2.4.4
- TensorFlow versione 1.14.0
- Zeppelin versione 0.9.0-SNAPSHOT
- Zookeeper versione 3.4.14
- Connettori e driver: DynamoDB Connector 4.14.0

 Note

Flink, Sqoop, Pig e Mahout non sono disponibili in Amazon EMR versione 6.0.0.

Nuove funzionalità di

- Supporto runtime docker YARN – Applicazioni YARN, come i lavori Spark, ora possono essere eseguite nel contesto di un container Docker. Ciò consente di definire facilmente le dipendenze in un'immagine Docker senza la necessità di installare librerie personalizzate nel cluster Amazon EMR. Per ulteriori informazioni, consulta [Configurazione dell'integrazione Docker](#) ed [Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.0.0](#).
- Supporto LLAP Hive - Hive ora supporta la modalità di esecuzione LLAP per migliorare le prestazioni delle query. Per ulteriori informazioni, consulta la sezione relativa all'[utilizzo di LLAP Hive](#).

Modifiche, miglioramenti e problemi risolti

- Questa versione risolve i problemi di dimensionamento di Amazon EMR nei casi in cui non riesce ad aumentare/ridurre correttamente un cluster o causa errori dell'applicazione.
- È stato risolto il problema per cui le richieste di dimensionamento avevano esito negativo per i cluster di grandi dimensioni e ad alto utilizzo quando i daemon su cluster di Amazon EMR eseguivano attività di controllo dell'integrità, come la raccolta dello stato del nodo YARN e dello stato del nodo HDFS. Ciò si verificava perché i daemon su cluster non erano in grado di comunicare i dati sullo stato di integrità di un nodo ai componenti interni di Amazon EMR.
- I daemon su cluster di EMR sono stati migliorati per monitorare correttamente gli stati dei nodi quando gli indirizzi IP vengono riutilizzati per migliorare l'affidabilità durante le operazioni di dimensionamento.
- [SPARK-29683](#). È stato risolto il problema per cui si verificavano errori di processo durante la riduzione del cluster poiché Spark presupponeva che tutti i nodi disponibili fossero negati.
- [YARN-9011](#). È stato risolto il problema per cui si verificavano errori di processo a causa di una race condition nella disattivazione di YARN quando il cluster tentava di aumentare o ridursi.
- È stato risolto il problema relativo agli errori di fase o processo durante il dimensionamento del cluster garantendo che gli stati dei nodi siano sempre coerenti tra i daemon su cluster di Amazon EMR e YARN/HDFS.

- È stato risolto il problema a causa del quale le operazioni del cluster, come la riduzione e l'invio di fasi, non riuscivano per i cluster Amazon EMR abilitati con autenticazione Kerberos. Questo era dovuto al fatto che il daemon su cluster di Amazon EMR non rinnovava il ticket di Kerberos, necessario per comunicare in modo sicuro con HDFS/YARN in esecuzione sul nodo primario.
- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Amazon Linux
 - Amazon Linux 2 è il sistema operativo per la serie EMR 6.x.
 - `systemd` viene utilizzato per la gestione dei servizi al posto di `upstart` utilizzato in Amazon Linux 1.
- Java Development Kit (JDK)
 - Corretto JDK 8 è il JDK predefinito per la serie delle versioni EMR 6.x.
- Scala
 - Scala 2.12 è usato con Apache Spark e Apache Livy.
- Python 3
 - Python 3 è ora la versione predefinita di Python in EMR.
- Etichette nodo YARN
 - A partire dalla serie di rilascio Amazon EMR 6.x, la funzione etichette nodo YARN è disabilitata per impostazione predefinita. Per impostazione predefinita, i processi master dell'applicazione possono essere eseguiti su entrambi i nodi di task e core. È possibile abilitare la caratteristica etichette nodo YARN configurando le seguenti proprietà: `yarn.node-labels.enabled` e `yarn.node-labels.am.default-node-label-expression`. Per ulteriori informazioni, consulta la sezione [Understanding Primary, Core, and Task Nodes](#) (Informazioni sui nodi primari, core e attività).

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` ed `emr-6.2.0` si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1,

5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione `ulimit` di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione `ulimit` durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'`ulimit` del controller dell'istanza su un massimo di 65536 file.

Impostazione di un `ulimit` esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un `ulimit` usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'`ulimit` del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
```

```
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- La shell interattiva di Spark, inclusi PySpark, SparkR e spark-shell, non supporta l'utilizzo di Docker con librerie aggiuntive.
- Per utilizzare Python 3 con Amazon EMR versione 6.0.0, è necessario aggiungere PATH a `yarn.nodemanager.env-whitelist`.
- La funzionalità Live Long and Process (LLAP) non è supportata quando si utilizza AWS Glue Data Catalog come metastore per Hive.
- Quando si utilizza Amazon EMR 6.0.0 con l'integrazione di Spark e Docker, è necessario configurare le istanze nel cluster con lo stesso tipo di istanza e la stessa quantità di volumi EBS per evitare errori durante l'invio di un processo Spark con il runtime di Docker.
- In Amazon EMR 6.0.0, la modalità di archiviazione di HBase su Amazon S3 è influenzata dal problema [HBASE-24286](#). Impossibile inizializzare il master HBase quando il cluster viene creato utilizzando dati S3 esistenti.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.


```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.30.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.30.1. Le modifiche sono relative alla versione 5.30.0.

Data del rilascio iniziale: 30 giugno 2020

Ultimo aggiornamento: 24 agosto 2020

Modifiche, miglioramenti e problemi risolti

- Le versioni più recenti di Amazon EMR risolvono il problema con un limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente di Amazon EMR. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono ora una correzione permanente con un'impostazione "Max open files" (Max. file aperti).
- Risolto il problema per cui il processo del controller di istanza generava un numero infinito di processi.
- Risolto un problema per cui Hue non era in grado di eseguire una query Hive, mostrando il messaggio "database is locked (database bloccato)" e impedendo l'esecuzione di query.
- Risolto un problema Spark per consentire l'esecuzione simultanea di più attività nel cluster EMR.
- Risolto un problema del notebook Jupyter che causava l'errore "too many open files (troppi file aperti)" nel server Jupyter.
- Corretto un problema relativo all'ora di inizio del cluster.

Nuove funzionalità di

- Le interfacce utente delle applicazioni persistenti Timeline Server di YARN e Tez sono disponibili con Amazon EMR versioni 6.x ed EMR versione 5.30.1 e successive. L'accesso ai collegamenti One-Click alla cronologia dell'applicazione persistente consente di accedere rapidamente alla cronologia dei processi senza configurare un proxy Web tramite una connessione SSH. I log dei cluster attivi e terminati sono disponibili per 30 giorni dopo la fine dell'applicazione. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce utente delle applicazioni persistenti](#) nella Guida alla gestione di Amazon EMR.
- Le API di esecuzione di EMR Notebooks sono disponibili per eseguire i notebook EMR tramite uno script o una riga di comando. La possibilità di avviare, arrestare, elencare e descrivere le esecuzioni dei notebook EMR senza la console AWS consente di controllare a livello di codice un notebook EMR. Utilizzando una cella del notebook parametrizzata, è possibile passare diversi valori di parametro a un notebook senza doverne creare una copia per ogni nuovo set di valori dei parametri. Consulta [Operazioni dell'API EMR](#). Per un codice di esempio, consulta [Comandi di esempio per eseguire EMR Notebooks a livello di codice](#).

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 ed emr-6.2.0 si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione ulimit inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione ulimit di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione ulimit inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione ulimit durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'ulimit del controller dell'istanza su un massimo di 65536 file.

Impostazione di un ulimit esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- [EMR Notebooks](#)

La caratteristica che consente di installare kernel e librerie Python aggiuntive sul nodo primario del cluster è disabilitata per impostazione predefinita in EMR versione 5.30.1. Per ulteriori informazioni su questa caratteristica, consulta la sezione [Installing Kernels and Python Libraries on a Cluster Primary Node](#) (Installazione di kernel e librerie Python su un nodo primario del cluster).

Per abilitare questa funzione, procedi come segue:

1. Assicurati che le policy di autorizzazione associate al ruolo di servizio per EMR Notebooks consentano l'operazione seguente:

```
elasticmapreduce:ListSteps
```

Per ulteriori informazioni, consultare il [Ruolo del servizio per EMR Notebooks](#).

2. Utilizza la AWS CLI per eseguire un passaggio sul cluster che imposta EMR Notebooks, come mostrato nell'esempio seguente. Sostituisci *us-east-1* con la Regione in cui risiede il cluster. Per ulteriori informazioni, consulta [Aggiunta di fasi a un cluster utilizzando la AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- Dimensionamento gestito

Le operazioni di dimensionamento gestito su cluster 5.30.0 e 5.30.1 senza Presto installato possono causare errori delle applicazioni o far sì che un gruppo di istanze o un parco istanze uniforme mantenga lo stato ARRESTED, in particolare quando un'operazione di dimensionamento verso il basso è seguita rapidamente da un'operazione di dimensionamento verso l'alto.

Come soluzione alternativa, scegli Presto come applicazione da installare quando crei un cluster con Amazon EMR rilasci 5.30.0 e 5.30.1, anche se il tuo processo non richiede Presto.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica

che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.
 - Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere `/` (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra `a` e `b` in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 5.30.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.30.0. Le modifiche sono relative alla versione 5.29.0.

Data del rilascio iniziale: 13 marzo 2020

Ultimo aggiornamento: 25 giugno 2020

Aggiornamenti

- Aggiornato AWS SDK for Java alla versione 1.11.759
- Aggiornamento di Amazon SageMaker Spark SDK alla versione 1.3.0
- Aggiornato EMR Record Server alla versione 1.6.0
- Flink aggiornato alla versione 1.10.0
- Aggiornato Ganglia alla versione 3.7.2
- HBase aggiornato alla versione 1.4.13
- Aggiornato Hudi alla versione 0.5.2-incubating
- Hue aggiornato alla versione 4.6.0
- Aggiornato JupyterHub alla versione 1.1.0
- Livy aggiornato alla versione 0.7.0-incubating
- Aggiornato Oozie alla versione 5.2.0
- Presto aggiornato alla versione 0.232
- Aggiornato Spark alla versione 2.4.5
- Connettori e driver aggiornati: connettore Amazon Glue 1.12.0; connettore Amazon Kinesis 3.5.0; connettore EMR DynamoDB 4.14.0

Nuove funzionalità di

- EMR Notebooks: se utilizzato con cluster EMR creati con 5.30.0, i kernel di EMR Notebooks vengono eseguiti sul cluster. Ciò migliora le prestazioni dei notebook e consente di installare e personalizzare i kernel. Puoi anche installare librerie Python sul nodo primario del cluster. Per ulteriori informazioni, consulta l'argomento relativo all'[installazione e l'utilizzo di kernel e librerie](#) nella Guida alla gestione di EMR.
- Dimensionamento gestito: con la versione 5.30.0 e successive di Amazon EMR, puoi abilitare il dimensionamento gestito da EMR per aumentare o diminuire automaticamente il numero di istanze

o unità nel cluster in base al carico di lavoro. Amazon EMR valuta continuamente i parametri dei cluster per prendere decisioni di dimensionamento che ottimizzano i cluster in termini di costi e velocità. Per maggiori informazioni, consulta [Dimensionamento delle risorse del cluster](#) nella Guida alla gestione di Amazon EMR.

- Crittografia dei file di log archiviati in Amazon S3: con Amazon EMR versione 5.30.0 e successive, puoi crittografare i file di log archiviati in Amazon S3 con una chiave gestita dal cliente AWS KMS. Per ulteriori informazioni, consulta [Crittografia dei file di log archiviati in Amazon S3](#) nella Guida alla gestione di Amazon EMR.
- Supporto di Amazon Linux 2: nella versione EMR 5.30.0 e successive, EMR utilizza il sistema operativo Amazon Linux 2. Le nuove AMI personalizzate (Amazon Machine Image) devono essere basate sull'AMI Amazon Linux 2. Per ulteriori informazioni, consulta [Utilizzo di un AMI personalizzato](#).
- Scalabilità automatica Presto con tolleranza: i cluster EMR che usano 5.30.0 possono essere impostati con un periodo di timeout di scalabilità automatica che consente alle attività Presto di terminare l'esecuzione prima che il loro nodo venga disattivato. Per ulteriori informazioni, consulta [Utilizzo della scalabilità automatica Presto con disattivazione con tolleranza](#).
- Creazione di parchi istanze con nuova opzione di strategia di allocazione: una nuova opzione di strategia di allocazione è disponibile in EMR versione 5.12.1 e successive. Offre un provisioning dei cluster più rapido, un'allocazione di istanze Spot più accurata e una minore interruzione delle istanze Spot. Sono necessari aggiornamenti ai ruoli di servizio EMR non predefiniti. Consulta [Configurazione di parchi istanze](#).
- Comandi `sudo systemctl stop` e `sudo systemctl start`: nella versione EMR 5.30.0 e successive, le quali usano il sistema operativo Amazon Linux 2 OS, EMR utilizza i comandi `sudo systemctl stop` e `sudo systemctl start` per riavviare i servizi. Per ulteriori informazioni, consulta [Come riavviare un servizio in Amazon EMR?](#).

Modifiche, miglioramenti e problemi risolti

- EMR versione 5.30.0 non installa Ganglia per impostazione predefinita. È possibile selezionare esplicitamente Ganglia da installare quando si crea un cluster.
- Ottimizzazione delle prestazioni di Spark.
- Ottimizzazione delle prestazioni di Presto.
- Python 3 è l'impostazione predefinita per Amazon EMR 5.30.0 e versioni successive.
- Il gruppo di sicurezza gestito predefinito per l'accesso al servizio nelle sottoreti private è stato aggiornato con nuove regole. Se si utilizza un gruppo di sicurezza personalizzato per l'accesso

al servizio, è necessario includere le stesse regole del gruppo di sicurezza gestito predefinito. Per ulteriori informazioni, consulta [Gruppo di sicurezza gestito da Amazon EMR per l'accesso al servizio \(sottoreti private\)](#). Se si utilizza un ruolo di servizio personalizzato per Amazon EMR, è necessario concedere l'autorizzazione a `ec2:describeSecurityGroups` per consentire a EMR di convalidare se i gruppi di sicurezza sono stati creati correttamente. Se si utilizza `EMR_DefaultRole`, questa autorizzazione è già inclusa nella policy gestita di default.

Problemi noti

- Riduci il limite "Max open files" (Max. file aperti) sulla versione di AL2 meno recente [corretto nelle versioni più recenti]. Rilasci di Amazon EMR: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` ed `emr-6.2.0` si basano sulle versioni precedenti di Amazon Linux 2 (AL2), le quali hanno un'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" quando i cluster Amazon EMR vengono creati con l'AMI predefinita. Le versioni di Amazon EMR `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` e successive includono una correzione permanente con un'impostazione "Max open files" (Max. file aperti). Le versioni con il limite minimo del file aperto causano l'errore "Too many open files" (Troppi file aperti) durante l'invio del processo Spark. Nelle versioni interessate, l'AMI predefinita di Amazon EMR ha un'impostazione `ulimit` di default di 4096 per "Max open files" (Max. file aperti), che è inferiore al limite di file 65536 nell'ultima AMI di Amazon Linux 2. L'impostazione `ulimit` inferiore per "Max open files (Max. file aperti)" causa il fallimento del processo Spark quando il driver e l'executor Spark tentano di aprire più di 4096 file. Per risolvere il problema, Amazon EMR dispone di uno script dell'operazione bootstrap (BA) che regola l'impostazione `ulimit` durante la creazione del cluster.

Se utilizzi una versione di Amazon EMR meno recente che non ha la soluzione permanente per questo problema, la seguente soluzione alternativa consente di impostare esplicitamente l'`ulimit` del controller dell'istanza su un massimo di 65536 file.

Impostazione di un `ulimit` esplicito dalla riga di comando

1. Modifica `/etc/systemd/system/instance-controller.service` per aggiungere i seguenti parametri alla sezione Servizio.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Riavvio di InstanceController


```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Impostazione di un ulimit usando l'operazione di bootstrap (BA)

È inoltre possibile utilizzare uno script dell'operazione di bootstrap (BA) per configurare l'ulimit del controller di istanza a 65536 file durante la creazione del cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Dimensionamento gestito

Le operazioni di dimensionamento gestito su cluster 5.30.0 e 5.30.1 senza Presto installato possono causare errori delle applicazioni o far sì che un gruppo di istanze o un parco istanze uniforme mantenga lo stato ARRESTED, in particolare quando un'operazione di dimensionamento verso il basso è seguita rapidamente da un'operazione di dimensionamento verso l'alto.

Come soluzione alternativa, scegli Presto come applicazione da installare quando crei un cluster con Amazon EMR rilasci 5.30.0 e 5.30.1, anche se il tuo processo non richiede Presto.

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

- Il motore di database predefinito per Hue 4.6.0 è SQLite, che causa problemi quando si tenta di utilizzare Hue con un database esterno. Per risolvere questo problema, imposta `engine` nella tua classificazione di configurazione `hue.ini` su `mysql`. Questo problema è stato risolto nella versione 5.30.1 di Amazon EMR.
- Se utilizzi Spark con la formattazione della posizione delle partizioni Hive per leggere i dati in Amazon S3 ed esegui Spark nei rilasci di Amazon EMR da 5.30.0 a 5.36.0 e da 6.2.0 a 6.9.0, potresti riscontrare un problema che impedisce al cluster di leggere correttamente i dati. Ciò può accadere se le partizioni presentano tutte le seguenti caratteristiche:
 - Due o più partizioni vengono scansionate dalla stessa tabella.
 - Almeno un percorso di directory di partizione è il prefisso di almeno un altro percorso della directory di partizione, ad esempio `s3://bucket/table/p=a` è un prefisso di `s3://bucket/table/p=a b`.

- Il primo carattere che segue il prefisso nell'altra directory di partizione ha un valore UTF-8 inferiore al carattere / (U+002F). Ad esempio, rientra in questa categoria il carattere dello spazio (U+0020) che compare tra a e b in `s3://bucket/table/p=a b`. Tieni presente che esistono altri 14 caratteri non di controllo: `!"#$%&'()*+,-.` Per ulteriori informazioni, consulta la sezione [UTF-8 encoding table and Unicode characters](#) (Tabella di codifica UTF-8 e caratteri Unicode).

Per ovviare a questo problema, imposta la configurazione di `spark.sql.sources.fastS3PartitionDiscovery.enabled` su `false` nella classificazione di `spark-defaults`.

Rilascio 5.29.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.29.0. Le modifiche sono relative alla versione 5.28.1.

Data del rilascio iniziale: 17 gennaio 2020

Aggiornamenti

- Aggiornato AWS SDK for Java alla versione 1.11.682
- Hive aggiornato alla versione 2.3.6
- Flink aggiornato alla versione 1.9.1
- EmrFS aggiornato alla versione 2.38.0
- EMR DynamoDB Connector aggiornato alla versione 4.13.0

Modifiche, miglioramenti e problemi risolti

- Spark
 - Ottimizzazione delle prestazioni di Spark.
- EMRFS
 - La Guida alla gestione aggiorna le impostazioni predefinite di `emrfs-site.xml` per una visualizzazione coerente.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.28.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.28.1. Le modifiche sono relative alla versione 5.28.0.

Data del rilascio iniziale: 10 gennaio 2020

Modifiche, miglioramenti e problemi risolti

- Spark
 - Risolti i problemi di compatibilità di Spark.
- Parametri di CloudWatch

- È stata corretta la pubblicazione dei parametri di Amazon CloudWatch su un cluster EMR con più nodi primari.
- Messaggio di log disattivato
 - Messaggio di falso log disabilitato, "[...] utilizzando la vecchia versione (<4.5.8) del client http Apache".

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.28.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.28.0. Le modifiche sono relative alla versione 5.27.0.

Data del rilascio iniziale: 12 novembre 2019

Aggiornamenti

- Flink aggiornato alla versione 1.9.0
- Hive aggiornato alla versione 2.3.6
- MXNet aggiornato alla versione 1.5.1
- Phoenix aggiornato alla versione 4.14.3
- Presto aggiornato alla versione 0.227
- Zeppelin aggiornato alla versione 0.8.2

Nuove funzionalità di

- [Apache Hudi](#) è ora disponibile per Amazon EMR per l'installazione durante la creazione di un cluster. Per ulteriori informazioni, consulta [Hudi](#).
- (25 novembre 2019) È ora possibile scegliere di eseguire più fasi in parallelo per migliorare l'utilizzo dei cluster e risparmiare sui costi. È anche possibile annullare sia le fasi in attesa che quelle in esecuzione. Per ulteriori informazioni, consultare [Utilizzo di fasi mediante l'AWS CLI e la console](#).
- (3 dicembre 2019) Adesso è possibile creare ed eseguire cluster EMR su AWS Outposts. AWS Outposts abilita i servizi nativi di, l'infrastruttura e i modelli operativi di AWS nelle strutture in locale. Negli ambienti AWS Outposts è possibile utilizzare le stesse API, strumenti e infrastruttura AWS utilizzati in AWS Cloud. Per ulteriori informazioni, consulta [Cluster EMR su AWS Outposts](#).
- (11 marzo 2020) A partire dalla versione Amazon EMR 5.28.0, è possibile creare ed eseguire i cluster Amazon EMR in una sottorete di AWS Local Zones come estensione logica di una Regione AWS che supporta le Local Zones. Una Local Zone consente ad Amazon EMR di individuare le funzionalità e un subset di servizi AWS, come i servizi di calcolo e archiviazione, per essere il più possibile vicina agli utenti, fornendo accesso a latenza molto bassa alle applicazioni in esecuzione localmente. Per un elenco delle Local Zones disponibili, consulta [AWS Local Zones](#). Per informazioni sull'accesso disponibile per le Local Zones AWS, consulta [Regioni, zone di disponibilità e Local Zones](#).

Attualmente, Local Zones non supporta Amazon EMR Notebooks e non supporta connessioni effettuate direttamente ad Amazon EMR utilizzando l'interfaccia endpoint VPC (AWS PrivateLink).

Modifiche, miglioramenti e problemi risolti

- Ampliato il supporto delle applicazioni per i cluster a elevata disponibilità
 - Per ulteriori informazioni, consulta la sezione [Applicazioni supportate in un cluster EMR con più nodi primari](#) nella Guida alla gestione di Amazon EMR.
- Spark
 - Ottimizzazione delle prestazioni
- Hive
 - Ottimizzazione delle prestazioni
- Presto
 - Ottimizzazione delle prestazioni

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

 Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Release 5.27.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.27.0. Le modifiche sono relative alla versione 5.26.0.

Data del rilascio iniziale: 23 settembre 2019

Aggiornamenti

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- TensorFlow 1.14.0
- Connettori e driver:
 - DynamoDB Connector 4.12.0

Nuove funzionalità di

- (24 ottobre 2019) Le seguenti nuove caratteristiche in EMR Notebooks sono disponibili con tutte le versioni di Amazon EMR
 - Ora è possibile associare repository Git ai notebook EMR per archiviare i notebook in un ambiente controllato dalla versione. È possibile condividere il codice con i peer e riutilizzare i notebook Jupyter esistenti tramite repository Git remoti. Per ulteriori informazioni, consulta [Associazione di repository Git ad Amazon EMR Notebooks](#) nella Guida alla gestione di Amazon EMR.
 - L'[utility nbdime](#) è ora disponibile in EMR Notebooks per semplificare il confronto e la fusione di notebook.

- Adesso, EMR Notebooks supporta JupyterLab. JupyterLab è un ambiente di sviluppo interattivo basato sul Web completamente compatibile con i notebook Jupyter. Ora è possibile scegliere di aprire il notebook nell'editor del notebook Jupyter o JupyterLab.
- (30 ottobre 2019) Con Amazon EMR versione 5.25.0 e successive, è possibile connettersi all'interfaccia utente di Spark History Server dalla pagina Summary (Riepilogo) del cluster o dalla scheda Application history (Cronologia applicazioni) nella console. Anziché configurare un proxy Web tramite una connessione SSH, puoi accedere rapidamente all'interfaccia utente del server della cronologia Spark per visualizzare i parametri dell'applicazione e accedere ai file di log pertinenti per cluster attivi e terminati. Per ulteriori informazioni, consulta [Accesso fuori cluster alle interfacce utente dell'applicazione persistente](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Cluster Amazon EMR con più nodi primari
 - È possibile installare ed eseguire Flink su un cluster Amazon EMR con più nodi primari. Per ulteriori informazioni, consulta [Applicazioni e funzionalità supportate](#).
 - È possibile configurare la crittografia trasparente HDFS su un cluster Amazon EMR con più nodi primari. Per maggiori informazioni, consulta la sezione [Crittografia trasparente HDFS su cluster EMR con più nodi primari](#).
 - È ora possibile modificare la configurazione delle applicazioni in esecuzione su un cluster Amazon EMR con più nodi primari. Per ulteriori informazioni, consulta [Specificazione di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).
- Connettore Amazon EMR-DynamoDB
 - Adesso, il connettore Amazon EMR-DynamoDB supporta i seguenti tipi di dati DynamoDB: booleano, elenco, mappa, elemento, null. Per ulteriori informazioni, consulta [Impostazione di una tabella Hive per l'esecuzione dei comandi Hive](#).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il

problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.26.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.26.0. Le modifiche sono relative alla versione 5.25.0.

Data del rilascio iniziale: 8 agosto 2019

Ultimo aggiornamento: 19 agosto 2019

Aggiornamenti

- AWS SDK for Java 1.11.595
- HBase 1.4.10
- Phoenix 4.14.2
- Connettori e driver:
 - DynamoDB Connector 4.11.0

- MariaDB Connector 2.4.2
- Driver JDBC Amazon Redshift 1.2.32.1056

Nuove funzionalità di

- (Beta) Con Amazon EMR 5.26.0 puoi avviare un cluster che si integra con Lake Formation. L'integrazione fornisce un accesso granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Inoltre, consente l'accesso federato single sign-on a EMR Notebooks o Apache Zeppelin da un sistema di identità aziendale. Per ulteriori informazioni, consulta [Integrazione di Amazon EMR con AWS Lake Formation \(Beta\)](#).
- (19 agosto 2019) Il blocco dell'accesso pubblico in Amazon EMR è ora disponibile con tutti i rilasci di Amazon EMR che supportano i gruppi di sicurezza. Il blocco dell'accesso pubblico è un'impostazione a livello di account applicata a ogni Regione AWS. Il blocco degli accessi pubblici impedisce l'avvio di un cluster quando per un gruppo di sicurezza a esso associato è impostata una regola che consente il traffico in entrata da IPv4 0.0.0.0/0 o IPv6 ::/0 (accesso pubblico) su una porta, a meno che la porta non sia specificata come eccezione. La porta 22 è un'eccezione per impostazione predefinita. Per ulteriori informazioni, consulta [Utilizzo del blocco dell'accesso pubblico in Amazon EMR](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Notebook EMR
 - Con EMR 5.26.0 e versioni successive, EMR Notebooks supporta librerie Python con ambito notebook oltre alle librerie Python predefinite. È possibile installare librerie con ambito notebook dall'editor di notebook senza dover ricreare un cluster o ricollegare un notebook a un cluster. Le librerie con ambito notebook vengono create in un ambiente Python virtuale in modo da essere applicate esclusivamente alla sessione di notebook corrente. Ciò consente di isolare le dipendenze dei notebook. Per ulteriori informazioni, consulta [Utilizzo di librerie con ambito notebook](#) nella Guida alla gestione di Amazon EMR.
- EMRFS
 - È possibile abilitare una funzione di verifica ETag (Beta) impostando `fs.s3.consistent.metadata.etag.verification.enabled` su `true`. Con questa caratteristica, EMRFS utilizza gli ETag Amazon S3 per verificare che gli oggetti letti siano la versione più recente disponibile. Questa caratteristica è utile per i casi d'uso read-after-update in cui i file in Amazon S3 vengono sovrascritti mantenendo lo stesso nome. Questa caratteristica

di verifica ETag attualmente non funziona con S3 Select. Per ulteriori informazioni, consulta [Configurazione della visualizzazione coerente](#).

- Spark
 - Le seguenti ottimizzazioni sono ora abilitate per impostazione predefinita: eliminazione dinamica delle partizioni, DISTINCT prima di INTERSECT, miglioramenti nell'inferenza delle statistiche del piano SQL per JOIN seguita da query DISTINCT, appiattimento delle sottoquery scalari, riordino di join ottimizzato e join del filtro Bloom. Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).
 - Migliorata la generazione del codice della fase per Sort Merge Join.
 - Migliorato il frammento di query e il riutilizzo delle sottoquery.
 - Miglioramenti per assegnare preventivamente gli executor all'avvio di Spark.
 - I join del filtro Bloom non vengono più applicati quando il lato più piccolo del join include un suggerimento per la trasmissione.
- Tez
 - Risolto un problema con Tez. Ora l'interfaccia utente di Tez funziona su un cluster Amazon EMR con più nodi primari.

Problemi noti

- Le funzionalità migliorate di generazione del codice dell'intera fase per Sort Merge Join possono aumentare la pressione della memoria quando abilitate. Questa ottimizzazione migliora le prestazioni, ma può comportare tentativi multipli o errori di processo se `spark.yarn.executor.memoryOverheadFactor` non è regolato per fornire una memoria sufficiente. Per disabilitare questa funzionalità, imposta `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` su `false` (falso).
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.25.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.25.0. Le modifiche sono relative alla versione 5.24.1

Data del rilascio iniziale: 17 luglio 2019

Ultimo aggiornamento: 30 ottobre 2019

Amazon EMR 5.25.0

Aggiornamenti

- 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

Nuove funzionalità di

- (30 ottobre 2019) A partire da Amazon EMR versione 5.25.0, è possibile connettersi all'interfaccia utente di Spark History Server dalla pagina Summary (Riepilogo) del cluster o dalla scheda Application history (Cronologia applicazioni) nella console. Anziché configurare un proxy Web tramite una connessione SSH, puoi accedere rapidamente all'interfaccia utente del server della cronologia Spark per visualizzare i parametri dell'applicazione e accedere ai file di log pertinenti per cluster attivi e terminati. Per ulteriori informazioni, consulta [Accesso fuori cluster alle interfacce utente dell'applicazione persistente](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Spark
 - Migliorate le prestazioni di alcuni join utilizzando i filtri Bloom per pre-filtrare gli input. L'ottimizzazione è disattivata per impostazione predefinita e può essere abilitata impostando il parametro di configurazione Spark `spark.sql.bloomFilterJoin.enabled` su `true`.
 - Migliorate le prestazioni del raggruppamento per colonne di tipo stringa.
 - Migliorata la memoria dell'executor Spark predefinita e la configurazione dei nodi principali dei tipi di istanza R4 per i cluster senza HBase installato.
 - Risolto un problema precedente correlato alla caratteristica di eliminazione delle partizioni dinamiche in cui la tabella eliminata deve trovarsi sul lato sinistro del join.
 - Migliorata l'ottimizzazione di DISTINCT prima di INTERSECT da applicare a casi aggiuntivi che coinvolgono alias.
 - Migliore inferenza delle statistiche del piano SQL per JOIN seguita da query DISTINCT. Questo miglioramento è disattivato per impostazione predefinita e può essere abilitato impostando il parametro di configurazione Spark `spark.sql.statsImprovements.enabled` su `true`. Questa ottimizzazione è richiesta dalla funzionalità Distinct prima di Intersect e verrà abilitata automaticamente quando `spark.sql.optimizer.distinctBeforeIntersect.enabled` è impostato su `true`.
 - Ordine di join ottimizzato in base alle dimensioni della tabella e ai filtri. Questa ottimizzazione è disattivata per impostazione predefinita e può essere abilitata impostando il parametro di configurazione Spark `spark.sql.optimizer.sizeBasedJoinReorder.enabled` su `true`.

Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).

- EMRFS

- L'impostazione EMRFS, `fs.s3.buckets.create.enabled`, è ora disabilitata per impostazione predefinita. A seguito di alcuni test, abbiamo scoperto che la disattivazione di questa impostazione migliora le prestazioni e impedisce la creazione involontaria di bucket S3. Se l'applicazione si basa su questa funzionalità, è possibile attivarla impostando la proprietà `fs.s3.buckets.create.enabled` su `true` nella classificazione di configurazione `emrfs-site`. Per informazioni, consulta [Specifica di una configurazione durante la creazione di un cluster](#).
- Miglioramenti della crittografia del disco locale e della crittografia S3 nelle configurazioni di sicurezza (5 agosto 2019)
 - Separate le impostazioni di crittografia Amazon S3 dalle impostazioni di crittografia del disco locale nella configurazione della protezione.
 - Aggiunta un'opzione per abilitare la crittografia EBS con la versione 5.24.0 e successive. Selezionando questa opzione, il volume del dispositivo di root viene crittografato oltre ai volumi di archiviazione. Le versioni precedenti richiedevano l'uso di un'AMI personalizzata per crittografare il volume del dispositivo di root.
 - Per ulteriori informazioni, consulta [Opzioni di crittografia](#) nella Guida alla gestione di Amazon EMR.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Release 5.24.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.24.1. Le modifiche sono relative alla versione 5.24.0.

Data del rilascio iniziale: 26 giugno 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.


```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.24.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.24.0. Le modifiche sono relative alla versione 5.23.0.

Data del rilascio iniziale: 11 giugno 2019

Ultimo aggiornamento: 5 agosto 2019

Aggiornamenti

- 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
- Connettori e driver:
 - DynamoDB Connector 4.9.0
 - MariaDB Connector 2.4.1

- Driver JDBC Amazon Redshift 1.2.27.1051

Modifiche, miglioramenti e problemi risolti

- Spark
 - Aggiunta ottimizzazione per eliminare dinamicamente le partizioni. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.dynamicPartitionPruning.enabled` su `true`.
 - Migliorate le prestazioni delle query INTERSECT. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.optimizer.distinctBeforeIntersect.enabled` su `true`.
 - Aggiunta ottimizzazione per appiattare le sottoquery scalari con aggregati che utilizzano la stessa relazione. Per impostazione predefinita, l'ottimizzazione è disabilitata. Per abilitare questa opzione, occorre impostare il parametro di configurazione Spark `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` su `true`.
 - Migliorata la generazione del codice della fase.

Per ulteriori informazioni, consulta [Ottimizzazione delle prestazioni Spark](#).

- Miglioramenti della crittografia del disco locale e della crittografia S3 nelle configurazioni di sicurezza (5 agosto 2019)
 - Separate le impostazioni di crittografia Amazon S3 dalle impostazioni di crittografia del disco locale nella configurazione della protezione.
 - Aggiunta un'opzione per abilitare la crittografia EBS. Selezionando questa opzione, il volume del dispositivo di root viene crittografato oltre ai volumi di archiviazione. Le versioni precedenti richiedevano l'uso di un'AMI personalizzata per crittografare il volume del dispositivo di root.
 - Per ulteriori informazioni, consulta [Opzioni di crittografia](#) nella Guida alla gestione di Amazon EMR.

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali

la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.23.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.23.0. Le modifiche sono relative alla versione 5.22.0.

Data del rilascio iniziale: 1° aprile 2019

Ultimo aggiornamento: 30 aprile 2019

Aggiornamenti

- AWS SDK for Java 1.11.519

Nuove funzionalità di

- (30 aprile 2019) Con Amazon EMR rilascio 5.23.0 e successivi, è possibile avviare un cluster con tre nodi primari per supportare applicazioni a elevata disponibilità come YARN Resource Manager, HDFS NameNode, Spark, Hive e Ganglia. Con questa caratteristica, il nodo primario non rappresenta più un potenziale singolo punto di errore. Se uno dei nodi primari ha esito negativo, Amazon EMR esegue automaticamente il failover in un nodo primario in standby e sostituisce il nodo primario guasto con uno nuovo con le medesime operazioni di configurazione e di bootstrap. Per ulteriori informazioni, consulta la sezione [Plan and Configure Primary Nodes](#) (Pianificazione e configurazione dei nodi primari).

Problemi noti

- Interfaccia utente Tez (corretta in Amazon EMR rilascio 5.26.0)

L'interfaccia utente di Tez non funziona su un cluster EMR con più nodi primari.

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.22.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.22.0. Le modifiche sono relative alla versione 5.21.0.

⚠ Important

A partire da Amazon EMR rilascio 5.22.0, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste inviate ad Amazon S3. I rilasci precedenti di Amazon EMR impiegano AWS Signature Version 2 in alcuni casi, a meno che le note di rilascio indichino l'utilizzo esclusivo di Signature Version 4. Per ulteriori informazioni, consulta [Autenticazione delle richieste \(AWS Signature Version 4\)](#) e [Autenticazione delle richieste \(AWS Signature Version 2\)](#) nella Guida per gli sviluppatori di Amazon Simple Storage.

Data del rilascio iniziale: 20 marzo 2019

Aggiornamenti

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Connettori e driver:
 - DynamoDB Connector 4.8.0
 - MariaDB Connector 2.2.6
 - Driver JDBC Amazon Redshift 1.2.20.1043

Nuove funzionalità di

- Modificata la configurazione EBS predefinita per i tipi di istanza EC2 con archiviazione solo EBS. Quando crei un cluster utilizzando Amazon EMR rilascio 5.22.0 e successivi, la quantità di archiviazione EBS predefinita aumenta in base alle dimensioni dell'istanza. Inoltre, l'archiviazione aumentata viene divisa su più volumi, offrendo migliori prestazioni IOPS. Se desideri utilizzare una configurazione dello storage delle istanze EBS diversa, puoi specificarla al momento della creazione di un cluster EMR o aggiungere nodi a un cluster esistente. Per ulteriori informazioni sulla quantità di archiviazione e sul numero di volumi allocati per impostazione predefinita per ciascun tipo di istanza, consulta [Archiviazione EBS di default per istanze](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Spark
 - Introdotta una nuova proprietà di configurazione per Spark su YARN, `spark.yarn.executor.memoryOverheadFactor`. Il valore di questa proprietà è un fattore di scala che imposta il valore del sovraccarico della memoria su una percentuale di memoria dell'executor, con un minimo di 384 MB. Se l'overhead della memoria è impostato esplicitamente utilizzando `spark.yarn.executor.memoryOverhead`, questa proprietà non ha effetto. Il valore predefinito è `0.1875`, che rappresenta il 18,75%. Questo valore predefinito per Amazon EMR lascia più spazio nei container YARN per l'overhead della memoria dell'executor rispetto al 10% predefinito impostato internamente da Spark. L'impostazione predefinita di Amazon EMR del 18,75% ha mostrato un minor numero di errori relativi alla memoria nei benchmark TPC-DS.
 - Backport di [SPARK-26316](#) per migliorare le prestazioni.
- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file `keytab` si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.21.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.21.1. Le modifiche sono relative alla versione 5.21.0.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Problemi noti

- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle

impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.21.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.21.0. Le modifiche sono relative alla versione 5.20.0.

Data del rilascio iniziale: 18 febbraio 2019

Ultimo aggiornamento: 3 aprile 2019

Aggiornamenti

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Nuove funzionalità di

- (3 aprile 2019) Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console Amazon EMR, la AWS Command Line Interface (AWS CLI) o il kit SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Modifiche, miglioramenti e problemi risolti

- Zeppelin
 - Backport per [ZEPPELIN-3878](#).

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)

- Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Tez
 - Questo problema è stato risolto in Amazon EMR 5.22.0.

Quando ci si connette all'interfaccia utente Tez all'indirizzo `http://MasterDNS:8080/tez-ui` tramite una connessione SSH al nodo primario del cluster, viene visualizzato l'errore "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled (Operazione adattatore fallita: server Timeline (ATS) fuori portata. O è inattivo o CORS non è abilitato)" o le attività mostrano una dicitura N/A inaspettata.

Ciò è dovuto al fatto che l'interfaccia utente Tez effettua richieste al server Timeline YARN utilizzando `localhost` anziché il nome `host` del nodo primario. Come soluzione alternativa, è disponibile uno script da eseguire come operazione o fase di bootstrap. Lo script aggiorna il nome `host` nel file `Tez configs.env`. Per ulteriori informazioni e per la posizione dello script, consulta [Istruzioni di bootstrap](#).

- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente hadoop al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente hadoop.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.20.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.20.0. Le modifiche sono relative alla versione 5.19.0.

Data del rilascio iniziale: 18 dicembre 2018

Ultimo aggiornamento: 22 gennaio 2019

Aggiornamenti

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

Nuove funzionalità di

- (22 gennaio 2019) Kerberos in Amazon EMR è stato migliorato per supportare l'autenticazione delle entità principali da un KDC esterno. In questo modo, la gestione dei principali viene centralizzata perché più cluster possono condividere un unico server KDC esterno. Inoltre, il server KDC esterno può avere un trust tra realm con un dominio Active Directory. In questo modo, tutti i cluster possono autenticare i principali da Active Directory. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione Kerberos](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- AMI predefinita di Amazon Linux per Amazon EMR
 - Il pacchetto Python3 è stato aggiornato da python 3.4 a 3.6.
- Il committer ottimizzato S3 EMRFS
 - Il committer EMRFS ottimizzato per S3 è ora abilitato per impostazione predefinita, il che migliora le prestazioni di scrittura. Per ulteriori informazioni, consulta [Utilizzare il committer ottimizzato S3 EMRFS](#).
- Hive

- Backport per [HIVE-16686](#).
- Glue con Spark e Hive
 - In EMR 5.20.0 o versioni successive, l'eliminazione delle partizioni parallele viene abilitata automaticamente per Spark e Hive quando AWS Glue Data Catalog viene utilizzato come metastore. Questa modifica riduce significativamente i tempi di pianificazione delle query eseguendo più richieste in parallelo per recuperare le partizioni. Il numero totale di segmenti che è possibile eseguire contemporaneamente è compreso tra 1 e 10. Il valore di default è 5, che è l'impostazione consigliata. Puoi modificare questa impostazione specificando la proprietà `aws.glue.partition.num.segments` nella classificazione di configurazione `hive-site`. Se si verifica una limitazione, è possibile disattivare la funzione modificando il valore su 1. Per ulteriori informazioni, consulta [Struttura del segmento AWS Glue](#).

Problemi noti

- Hue (corretto in Amazon EMR rilascio 5.24.0)
 - Hue in esecuzione su Amazon EMR non supporta Solr. A partire da Amazon EMR rilascio 5.20.0, un problema di configurazione errata causa l'attivazione di Solr e un messaggio di errore innocuo simile al seguente:

```
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',))
```

Per evitare che venga visualizzato il messaggio di errore Solr:

1. Effettua la connessione alla linea di comando del nodo primario tramite SSH.
2. Aprire il file `hue.ini` con un editor di testo. Ad esempio:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Cerca il termine `appblacklist` e modifica la riga nel modo seguente:

```
appblacklist = search
```

4. Salva le modifiche e riavvia Hue come nell'esempio seguente:

```
sudo stop hue; sudo start hue
```

- Tez
 - Questo problema è stato risolto in Amazon EMR 5.22.0.

Quando ci si connette all'interfaccia utente Tez all'indirizzo `http://MasterDNS:8080/tez-ui` tramite una connessione SSH al nodo primario del cluster, viene visualizzato l'errore "Adapter operation failed - Timeline server (ATS) is out of reach. Either it is down, or CORS is not enabled (Operazione adattatore fallita: server Timeline (ATS) fuori portata. O è inattivo o CORS non è abilitato)" o le attività mostrano una dicitura N/A inaspettata.

Ciò è dovuto al fatto che l'interfaccia utente Tez effettua richieste al server Timeline YARN utilizzando `localhost` anziché il nome host del nodo primario. Come soluzione alternativa, è disponibile uno script da eseguire come operazione o fase di bootstrap. Lo script aggiorna il nome host nel file `Tez configs .env`. Per ulteriori informazioni e per la posizione dello script, consulta [Istruzioni di bootstrap](#).

- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.
- Problema noto nei cluster con più nodi primari e autenticazione Kerberos

Se esegui cluster con più nodi primari e autenticazione Kerberos nella versione 5.20.0 di Amazon EMR e successivi, è possibile che si verifichino problemi con le operazioni del cluster, quali la riduzione o l'invio di fasi, dopo che il cluster è rimasto in esecuzione per un certo periodo di tempo. Il periodo di tempo dipende dal periodo di validità del ticket Kerberos definito dall'utente. Il problema di dimensionamento verso il basso influisce sia sulle richieste di scalabilità automatica che sulle richieste esplicite di dimensionamento verso il basso inviate dall'utente. Possono essere influenzate anche le operazioni cluster aggiuntive.

Soluzione alternativa:

- Abilita SSH come utente `hadoop` al primo nodo primario del cluster EMR con più nodi primari.
- Esegui il comando seguente per rinnovare il ticket Kerberos per l'utente `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

In genere, il file keytab si trova in `/etc/hadoop.keytab` e l'entità principale si presenta nella forma di `hadoop/<hostname>@<REALM>`.

Note

Questa soluzione alternativa sarà valida per il periodo di tempo in cui il ticket Kerberos è valido. La durata è di 10 ore per impostazione predefinita, ma può essere configurata dalle impostazioni Kerberos. È necessario eseguire nuovamente il comando sopra indicato una volta scaduto il ticket Kerberos.

Rilascio 5.19.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.19.0. Le modifiche sono relative alla versione 5.18.0.

Data del rilascio iniziale: 7 novembre 2018

Ultimo aggiornamento: 19 novembre 2018

Aggiornamenti

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- ZooKeeper 3.4.13
- AWS SDK for Java 1.11.433

Nuove funzionalità di

- (19 novembre 2018) EMR Notebooks è un ambiente gestito basato su Jupyter Notebook. Supporta i kernel Spark magic per PySpark, Spark SQL, Spark, R e Scala. EMR Notebooks può essere utilizzato con cluster creati utilizzando Amazon EMR rilascio 5.18.0 e successivi. Per ulteriori informazioni, consulta [Utilizzo di EMR Notebooks](#) nella Guida alla gestione di Amazon EMR.

- Il committer EMRFS ottimizzato per S3 è disponibile durante la scrittura di file Parquet tramite Spark ed EMRFS. Questo committer migliora le prestazioni di scrittura. Per ulteriori informazioni, consulta [Utilizzare il committer ottimizzato S3 EMRFS](#).

Modifiche, miglioramenti e problemi risolti

- YARN
 - Modificata la logica che limita il processo master dell'applicazione all'esecuzione sui nodi principali. Questa funzionalità ora utilizza la caratteristica e le proprietà delle etichette dei nodi YARN nelle classificazioni di configurazione `yarn-site` e `capacity-scheduler`. Per informazioni, consultare, <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- AMI predefinita di Amazon Linux per Amazon EMR
 - `ruby18`, `php56` e `gcc48` non vengono più installati per impostazione predefinita. Se lo si desidera, è possibile installarli tramite `yum`.
 - Il Ruby Gem `aws-java-sdk` non è più installato per impostazione predefinita. Se lo si desidera, è possibile installarlo tramite `gem install aws-sdk`. Possono anche essere installati componenti specifici. Ad esempio, `gem install aws-sdk-s3`.

Problemi noti

- EMR Notebooks: in alcune circostanze, con più editor di notebook aperti, l'editor del notebook potrebbe non essere in grado di connettersi al cluster. In questo caso, cancella i cookie del browser e riapri gli editor dei notebook.
- Parametro `ContainerPending` e scalabilità automatica di CloudWatch: (corretto in 5.20.0) Amazon EMR può emettere un valore negativo per `ContainerPending`. Se `ContainerPending` viene utilizzato in una regola di scalabilità automatica, quest'ultima avrà un comportamento imprevisto. Non utilizzare `ContainerPending` con la scalabilità automatica.
- In Amazon EMR versione 5.19.0, 5.20.0 e 5.21.0, le etichette dei nodi YARN sono memorizzate in una directory HDFS. In alcune situazioni, questo porta a ritardi di avvio del nodo principale e quindi causa il timeout del cluster e l'errore di avvio. A partire da Amazon EMR 5.22.0, questo problema è stato risolto. Le etichette dei nodi YARN sono memorizzate sul disco locale di ciascun nodo cluster, evitando dipendenze da HDFS.

Rilascio 5.18.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.18.0. Le modifiche sono relative alla versione 5.17.0.

Data del rilascio iniziale: 24 ottobre 2018

Aggiornamenti

- Flink 1.6.0
- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Nuove funzionalità di

- A partire da Amazon EMR rilascio 5.18.0, è possibile utilizzare i repository degli artefatti di Amazon EMR per creare il codice di processo secondo le versioni esatte di librerie e dipendenze disponibili con rilasci Amazon EMR specifici. Per ulteriori informazioni, consulta [Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR](#).

Modifiche, miglioramenti e problemi risolti

- Hive
 - Aggiunto il supporto per S3 Select. Per ulteriori informazioni, consulta [Utilizzo di S3 Select con Hive per migliorare le prestazioni](#).
- Presto
 - Aggiunto il supporto per [S3 Select](#) Pushdown. Per ulteriori informazioni, consulta [Uso di pushdown S3 Select con Presto per migliorare le prestazioni](#).
- Spark
 - La configurazione predefinita log4j per Spark è stata modificata per eseguire i registri del container ogni ora per i processi di streaming Spark. Ciò consente di impedire l'eliminazione dei registri per i processi Spark Streaming con esecuzione prolungata.

Rilascio 5.17.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.17.1. Le modifiche sono relative alla versione 5.17.0.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Rilascio 5.17.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.17.0. Le modifiche sono relative alla versione 5.16.0.

Data del rilascio iniziale: 30 agosto 2018

Aggiornamenti

- Flink 1.5.2
- HBase 1.4.6
- Presto 0.206

Nuove funzionalità di

- Aggiunto il supporto per Tensorflow. Per ulteriori informazioni, consulta [TensorFlow](#).

Modifiche, miglioramenti e problemi risolti

- JupyterHub
 - Aggiunto il supporto per la persistenza notebook in Amazon S3. Per ulteriori informazioni, consulta [Configurazione della persistenza per i notebook in Amazon S3](#).
- Spark
 - Aggiunto il supporto per [S3 Select](#). Per ulteriori informazioni, consulta [Usare S3 Select con Spark per migliorare le prestazioni delle query](#).

- Risolti i problemi con i parametri Cloudwatch e la funzione di scalabilità automatica in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0.

Problemi noti

- Quando si crea un cluster che utilizza Kerberos con Livy installato, Livy ha esito negativo e genera un errore a indicare che l'autenticazione semplice non è abilitata. Il riavvio del server Livy risolve il problema. Per ovviare al problema, è possibile aggiungere una fase durante la creazione del cluster che esegue `sudo restart livy-server` sul nodo primario.
- Se si utilizza un'AMI Amazon Linux personalizzata basata su un'AMI Amazon Linux con una data di creazione 2018-08-11, il server Oozie non si avvia. Se si utilizza Oozie, è possibile creare un'AMI personalizzata basata su un ID AMI Amazon Linux con una data di creazione diversa. Puoi utilizzare il comando AWS CLI seguente per restituire un elenco di ID immagine per tutte le AMI Amazon Linux HVM con una versione 2018.03, insieme alla data di rilascio, in modo da poter scegliere un'AMI Linux appropriata come base. Sostituisci `MyRegion` con l'identificatore della Regione, ad esempio `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
```

Rilascio 5.16.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.16.0. Le modifiche sono relative alla versione 5.15.0.

Data del rilascio iniziale: 19 luglio 2018

Aggiornamenti

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0
- Phoenix 4.14.0
- Presto 0.203

- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA 9.2
- Driver JDBC Redshift 1.2.15.1025

Modifiche, miglioramenti e problemi risolti

- HBase
 - Backport per [HBASE-20723](#)
- Presto
 - Modifiche alla configurazione per supportare l'autenticazione LDAP. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione LDAP con Presto on Amazon EMR](#).
- Spark
 - Apache Spark versione 2.3.1, disponibile a partire da Amazon EMR rilascio 5.16.0, risolve [CVE-2018-8024](#) e [CVE-2018-1334](#). Si consiglia di eseguire la migrazione delle versioni precedenti di Spark a Spark 2.3.1 o versioni successive.

Problemi noti

- Questa versione non supporta i tipi di istanza c1.medium o m1.small. I cluster che utilizzano uno di questi tipi di istanza non vengono avviati. Come soluzione alternativa, specifica un tipo di istanza diverso o utilizza una versione di rilascio diversa.
- Quando si crea un cluster che utilizza Kerberos con Livy installato, Livy ha esito negativo e genera un errore a indicare che l'autenticazione semplice non è abilitata. Il riavvio del server Livy risolve il problema. Per ovviare al problema, è possibile aggiungere una fase durante la creazione del cluster che esegue `sudo restart livy-server` sul nodo primario.
- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Rilascio 5.15.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.15.0. Le modifiche sono relative alla versione 5.14.0.

Data del rilascio iniziale: 21 giugno 2018

Aggiornamenti

- Aggiornato a HBase 1.4.4
- Aggiornato a Hive 2.3.3
- Aggiornato a Hue 4.2.0
- Oozie aggiornato a 5.0.0
- Aggiornato a Zookeeper 3.4.12
- AWS SDK aggiornato alla versione 1.11.333

Modifiche, miglioramenti e problemi risolti

- Hive
 - Backport per [HIVE-18069](#).
- Hue
 - Hue aggiornato per autenticarsi correttamente con Livy quando Kerberos è abilitato. Livy è ora supportato quando si utilizza Kerberos con Amazon EMR.
- JupyterHub
 - JupyterHub aggiornato affinché Amazon EMR possa installare le librerie client LDAP per impostazione predefinita.
 - Corretto un errore nello script che genera certificati autofirmati.

Problemi noti

- Questa versione non supporta i tipi di istanza c1.medium o m1.small. I cluster che utilizzano uno di questi tipi di istanza non vengono avviati. Come soluzione alternativa, specifica un tipo di istanza diverso o utilizza una versione di rilascio diversa.
- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in

Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Rilascio 5.14.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.14.1. Le modifiche sono relative alla versione 5.14.0.

Data del rilascio iniziale: 17 ottobre 2018

Aggiornata l'AMI predefinita per Amazon EMR al fine di risolvere potenziali vulnerabilità di sicurezza.

Rilascio 5.14.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.14.0. Le modifiche sono relative alla versione 5.13.0.

Data del rilascio iniziale: 4 giugno 2018

Aggiornamenti

- Aggiornato ad Apache Flink 1.4.2
- Apache MXNet aggiornato alla versione 1.1.0
- Apache Sqoop aggiornato alla versione 1.4.7

Nuove funzionalità di

- Aggiunto il supporto JupyterHub. Per ulteriori informazioni, consulta [JupyterHub](#).

Modifiche, miglioramenti e problemi risolti

- EMRFS
 - La stringa userAgent nelle richieste ad Amazon S3 è stata aggiornata per contenere le informazioni sull'utente e sul gruppo dell'entità principale che richiama. Questa può essere utilizzata con i log AWS CloudTrail per un monitoraggio più completo delle richieste.
- HBase
 - Include [HBASE-20447](#), che risolve un problema che potrebbe causare errori di cache, in particolare in caso di Regioni divise.

- MXNet
 - Aggiunte librerie OpenCV.
- Spark
 - Quando Spark scrive i file Parquet in un percorso Amazon S3 utilizzando EMRFS, l'algoritmo FileOutputCommitter è stato aggiornato per utilizzare la versione 2 anziché la versione 1. In questo modo viene ridotta la quantità di ridenominazioni, con un conseguente miglioramento delle prestazioni delle applicazioni. Questa modifica non influisce su:
 - Applicazioni diverse da Spark.
 - Applicazioni che scrivono su altri file system, ad esempio HDFS (che utilizzano ancora la versione 1 di FileOutputCommitter).
 - Applicazioni che utilizzano altri formati di output, come testo o csv, che già utilizzano la scrittura diretta EMRFS.

Problemi noti

- JupyterHub
 - L'utilizzo delle classificazioni di configurazione per impostare JupyterHub e singoli notebook Jupyter quando si crea un cluster non è supportato. Modifica manualmente il file `jupyterhub_config.py` e i file `jupyter_notebook_config.py` per ogni utente. Per ulteriori informazioni, consulta [Configurazione di JupyterHub](#).
 - JupyterHub non si avvia sui cluster all'interno di una sottorete privata, generando il messaggio `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'`. Questo problema è causato da un errore nello script che genera certificati autofirmati. Utilizza la seguente soluzione alternativa per generare certificati autofirmati. Tutti i comandi vengono eseguiti mentre si è connessi al nodo primario.

1. Copia lo script di generazione del certificato dal container al nodo primario:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Utilizza un editor di testo per intervenire sulla riga 23 modificando il nome di host pubblico nel nome di host locale come illustrato di seguito:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Esegui lo script per generare certificati autofirmati:


```
sudo bash ./gen_self_signed_cert.sh
```

4. Sposta i file di certificato generati dallo script nella directory `/etc/jupyter/conf/`:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

È possibile `tail` il file `jupyter.log` per verificare che JupyterHub sia stato riavviato e stia restituendo un codice di risposta 200. Ad esempio:

```
tail -f /var/log/jupyter/jupyter.log
```

Questo dovrebbe restituire una risposta simile alla seguente:

```
# [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
```

- Dopo il riavvio del nodo primario o il riavvio del controller dell'istanza, i parametri CloudWatch non verranno acquisiti e la caratteristica di dimensionamento automatico non sarà disponibile in Amazon EMR versione 5.14.0, 5.15.0 o 5.16.0. Questo problema è stato risolto in Amazon EMR 5.17.0.

Rilascio 5.13.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.13.0. Le modifiche sono relative alla versione 5.12.0.

Aggiornamenti

- Spark aggiornato a 2.3.0
- HBase aggiornato a 1.4.2
- Presto aggiornato a 0.194
- Aggiornato AWS SDK for Java a 1.11.297

Modifiche, miglioramenti e problemi risolti

- Hive

- Backport per [HIVE-15436](#). API di Hive ottimizzate per restituire solo visualizzazioni.

Problemi noti

- Al momento MXNet non dispone di librerie OpenCV.

Rilascio 5.12.2

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.2. Le modifiche sono relative alla versione 5.12.1.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Rilascio 5.12.1

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.1. Le modifiche sono relative alla versione 5.12.0.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Rilascio 5.12.0

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.12.0. Le modifiche sono relative alla versione 5.11.1.

Aggiornamenti

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. Per ulteriori informazioni, consulta il [Log delle modifiche AWS SDK per Java](#) su GitHub
- Hadoop 2.7.3 ⇒ 2.8.3. Per ulteriori informazioni, consulta [Rilasci di Apache Hadoop](#).

- Flink 1.3.2 ⇒ 1.4.0. Per ulteriori informazioni, consulta l'[Annuncio del rilascio di Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Per ulteriori informazioni, consulta l'[Annuncio del rilascio di HBase](#).
- Hue 4.0.1 ⇒ 4.1.0. Per ulteriori informazioni, consulta le [Note di rilascio](#).
- MxNet 0.12.0 ⇒ 1.0.0. Per ulteriori informazioni, consulta il [Log delle modifiche MXNet](#) su GitHub.
- Presto 0.187 ⇒ 0.188. Per ulteriori informazioni, consulta le [Note di rilascio](#).

Modifiche, miglioramenti e problemi risolti

- Hadoop
 - La proprietà `yarn.resourcemanager.decommissioning.timeout` è stata modificata in `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. È possibile utilizzare questa proprietà per personalizzare il dimensionamento verso il basso del cluster. Per maggiori informazioni, consulta [Dimensionamento verso il basso del cluster](#) nella Guida alla gestione di Amazon EMR.
 - L'interfaccia CLI di Hadoop ha aggiunto l'opzione `-d` per il comando `cp` (copia), che specifica la copia diretta. Questo comando può essere utilizzato per evitare di creare un file `.COPYING` intermediario, rendendo più veloce la copia dei dati in Amazon S3. Per ulteriori informazioni, consulta [HADOOP-12384](#).
- Pig
 - È stato aggiunto `pig-env`, che semplifica la configurazione delle proprietà dell'ambiente Pig. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).
- Presto
 - Aggiunta la classificazione della configurazione `presto-connector-redshift`, che puoi utilizzare per configurare i valori nel file di configurazione `redshift.properties` di Presto. Per ulteriori informazioni, consulta [Connettore Redshift](#) nella documentazione di Presto e [Configurazione delle applicazioni](#).
 - Il supporto Presto per EMRFS è stato aggiunto ed è la configurazione predefinita. I rilasci precedenti di Amazon EMR utilizzavano `PrestoS3FileSystem`, che era l'unica opzione disponibile. Per ulteriori informazioni, consulta [Configurazione EMRFS e PrestoS3FileSystem](#).

Note

Se esegui una query sui dati sottostanti in Amazon S3 con Amazon EMR versione 5.12.0, possono verificarsi errori di Presto. Questo avviene perché Presto non riesce a raccogliere i valori di classificazione della configurazione da `emrfs-site.xml`. Come

soluzione alternativa, crea una sottodirectory `emrfs` in `usr/lib/presto/plugin/hive-hadoop2/` e crea un collegamento simbolico in `usr/lib/presto/plugin/hive-hadoop2/emrfs` sul file `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` esistente. Quindi riavvia il processo `presto-server` (`sudo presto-server stop` seguito da `sudo presto-server start`).

- Spark
 - Backport di [SPARK-22036: talvolta la moltiplicazione BigDecimal restituisce null](#).

Problemi noti

- MXNet non include le librerie OpenCV.
- SparkR non è disponibile per i cluster creati utilizzando un'AMI personalizzata perché R non è installato per impostazione predefinita sui nodi del cluster.

Rilascio 5.11.3

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.11.3. Le modifiche sono relative alla versione 5.11.2.

Data del rilascio iniziale: 18 luglio 2019

Modifiche, miglioramenti e problemi risolti

- Aggiornata l'AMI predefinita Amazon Linux per Amazon EMR per importanti aggiornamenti di sicurezza del kernel Linux, inclusi i problemi di negazione del servizio TCP SACK ([AWS-2019-005](#)).

Rilascio 5.11.2

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.11.2. Le modifiche sono relative alla versione 5.11.1.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Rilascio 5.11.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.11.1. Le modifiche sono relative al rilascio di Amazon EMR 5.11.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Problemi noti

- MXNet non include le librerie OpenCV.
- Per impostazione predefinita, Hive 2.3.2 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Rilascio 5.11.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.11.0. Le modifiche sono relative al rilascio di Amazon EMR 5.10.0.

Aggiornamenti

Le seguenti applicazioni e componenti sono stati aggiornati in questo rilascio per includere le seguenti versioni.

- Hive 2.3.2
- Spark 2.2.1
- SDK per Java: 1.11.238

Nuove funzionalità di

- Spark
 - Aggiunta l'impostazione `spark.decommissioning.timeout.threshold` che migliora il comportamento di disattivazione di Spark quando si utilizzano istanze Spot. Per ulteriori informazioni, consulta [Configurazione del comportamento di disattivazione di nodi](#).
 - Aggiunto a Spark il componente `aws-sagemaker-spark-sdk`, che installa Amazon SageMaker Spark e le dipendenze associate per l'integrazione di Spark con [Amazon SageMaker](#). Puoi utilizzare Amazon SageMaker Spark per configurare le pipeline di machine learning (ML) di Spark utilizzando le fasi di Amazon SageMaker. Per ulteriori informazioni, consulta il [File readme di SageMaker Spark](#) su GitHub e [Utilizzo di Apache Spark con Amazon SageMaker](#) nella Guida per lo sviluppatore di Amazon SageMaker.

Problemi noti

- MXNet non include le librerie OpenCV.
- Per impostazione predefinita, Hive 2.3.2 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Rilascio 5.10.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.10.0. Le modifiche sono relative al rilascio di Amazon EMR 5.9.0.

Aggiornamenti

Le seguenti applicazioni e componenti sono stati aggiornati in questo rilascio per includere le seguenti versioni.

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

Nuove funzionalità di

- Aggiunto supporto per l'autenticazione Kerberos. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione Kerberos](#) nella Guida alla gestione di Amazon EMR
- Aggiunto supporto per i ruoli IAM per richieste EMRFS ad Amazon S3. Per ulteriori informazioni, consulta [Configurazione dei ruoli IAM per le richieste EMRFS ad Amazon S3](#) nella Guida alla gestione di Amazon EMR.
- Aggiunto supporto per i tipi di istanza P2 e P3 basati su GPU. Per ulteriori informazioni, consulta [Istanze P2 di Amazon EC2](#) e [Istanze P3 di Amazon EC2](#). NVIDIA driver 384.81 e CUDA driver 9.0.176 sono installati su questi tipi di istanza per impostazione predefinita.
- Aggiunta del supporto per [Apache MXNet](#).

Modifiche, miglioramenti e problemi risolti

- Presto
 - Aggiunto supporto per l'utilizzo di AWS Glue Data Catalog come metastore predefinito di Hive. Per ulteriori informazioni, consulta [Utilizzo di Presto con AWS Glue Data Catalog](#).
 - Aggiunto supporto per [funzioni geospaziali](#).
 - Aggiunto supporto [spill to disk](#) per join.
 - Aggiunto supporto per il [connettore Redshift](#).
- Spark
 - Backport per [SPARK-20640](#), che rende il timeout rcp e i tentativi per i valori di registrazione shuffle configurabili utilizzando le `spark.shuffle.registration.timeout` e `spark.shuffle.registration.maxAttempts` proprietà.

- Backport per [SPARK-21549](#), che corregge un errore che avviene durante la scrittura di OutputFormat personalizzato su percorsi non-HDFS.
- Backport per [Hadoop-13270](#)
- Le librerie Numpy, Scipy e Matplotlib sono state rimosse dall'AMI di base di Amazon EMR. Se hai bisogno di queste librerie per tua applicazione, sono disponibili nella repository dell'applicazione, così potrai usare un'operazione di bootstrap per installarle su tutti i nodi usando `yum install`.
- L'AMI di base Amazon EMR non dispone più dei pacchetti RPM di applicazione inclusi, per cui i pacchetti RPM non sono più presenti sui nodi cluster. Adesso, le AMI personalizzate e l'AMI di base Amazon EMR fanno riferimento al repository dei pacchetti RPM in Amazon S3.
- A causa dell'introduzione della fatturazione al secondo in Amazon EC2, il valore predefinito Scale down behavior (Comportamento di dimensionamento verso il basso) è diventato Terminate at task completion (Termina al completamento dell'attività) anziché Terminate at instance hour (Termina all'ora dell'istanza). Per maggiori informazioni, consulta [Configurazione del ridimensionamento del cluster](#).

Problemi noti

- MXNet non include le librerie OpenCV.
- Per impostazione predefinita, Hive 2.3.1 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Rilascio 5.9.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.9.0. Le modifiche sono relative al rilascio di Amazon EMR 5.8.0.

Data del rilascio: 5 ottobre 2017

Ultimo aggiornamento della caratteristica: 12 ottobre 2017

Aggiornamenti

Le seguenti applicazioni e componenti sono stati aggiornati in questo rilascio per includere le seguenti versioni.

- AWS SDK for Java versione 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Nuove funzionalità di

- Aggiunto il supporto per Livy (versione 0.4.0-incubazione). Per ulteriori informazioni, consulta [Apache Livy](#).
- Aggiunto il supporto per Hue Notebook per Spark.
- Aggiunto supporto per le istanze Amazon EC2 serie i3 (12 ottobre 2017).

Modifiche, miglioramenti e problemi risolti

- Spark
 - Aggiunto un nuovo set di caratteristiche per garantire che Spark gestisca più elegantemente la terminazione dei nodi a causa di un ridimensionamento manuale o di una richiesta di policy di scalabilità automatica. Per ulteriori informazioni, consulta [Configurazione del comportamento di disattivazione di nodi](#).
 - Viene utilizzato SSL invece di 3DES per la crittografia in transito del servizio di trasferimento dei blocchi, migliorando così le prestazioni dei tipi di istanza Amazon EC2 con AES-NI.
 - Backport per [SPARK-21494](#).
- Zeppelin
 - Backport per [ZEPPELIN-2377](#).
- HBase
 - Aggiunta patch [HBASE-18533](#) che permette valori aggiuntivi per la configurazione HBase BucketCache utilizzando la classificazione della configurazione `hbase-site`.

- Hue
 - Aggiunto il supporto AWS Glue Data Catalog per l'editor delle query Hive su Hue.
 - Per impostazione predefinita, i superuser su Hue possono accedere a tutti i file a cui hanno accesso i ruoli IAM di Amazon EMR. Gli utenti appena creati non dispongono automaticamente delle autorizzazioni per accedere al filebrowser Amazon S3 e devono disporre delle autorizzazioni `filebrowser.s3_access` attivate per il proprio gruppo.
- Risolto un problema che causava la non accessibilità ai dati JSON sottostanti creati utilizzando AWS Glue Data Catalog.

Problemi noti

- L'avvio del cluster non va a buon fine quando tutte le applicazioni sono installate e la dimensione del volume di root Amazon EBS di default non viene modificata. Per risolvere il problema, utilizza il comando `aws emr create-cluster` da AWS CLI e specifica un parametro `--ebs-root-volume-size` più grande.
- Per impostazione predefinita, Hive 2.3.0 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio, se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella LOCATION, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

Rilascio 5.8.2

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.8.2. Le modifiche sono relative alla versione 5.8.1.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Rilascio 5.8.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.8.1. Le modifiche sono relative al rilascio di Amazon EMR 5.8.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Rilascio 5.8.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.8.0. Le modifiche sono relative al rilascio di Amazon EMR 5.7.0.

Data del rilascio iniziale: 10 agosto 2017

Ultimo aggiornamento della caratteristica: 25 settembre 2017

Aggiornamenti

Le seguenti applicazioni e componenti sono stati aggiornati in questo rilascio per includere le seguenti versioni:

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Per ulteriori informazioni, consulta [Release notes \(Note di rilascio\)](#) sul sito di Apache Hive.
- Spark 2.2.0. Per ulteriori informazioni, consulta [Release notes \(Note di rilascio\)](#) sul sito di Apache Spark.

Nuove funzionalità di

- Aggiunto supporto per la visualizzazione della cronologia dell'applicazione (25 settembre 2017). Per ulteriori informazioni, consulta [Visualizzazione della cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.

Modifiche, miglioramenti e problemi risolti

- Integrazione con Glue Data Catalog di AWS
 - Aggiunta funzionalità che permette a Hive e Spark SQL di utilizzare AWS Glue Data Catalog come archivio dei metadati di Hive. Per ulteriori informazioni, consulta [Utilizzo di AWS Glue Data Catalog come metastore Hive](#) e [Utilizzare AWS Glue Data Catalog come metastore Spark SQL](#).
- Aggiunto Application history (Cronologia dell'applicazione) ai dettagli del cluster, il che ti permette di visualizzare i dati cronologici delle applicazioni YARN e i dettagli aggiuntivi per le applicazioni Spark. Per ulteriori informazioni, consulta [Visualizza cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.
- Oozie
 - Backport per [OOZIE-2748](#).
- Hue
 - Backport per [HUE-5859](#)
- HBase
 - Aggiunta patch per esporre il tempo di avvio del server master HBase tramite le Java Management Extensions (JMX) utilizzando `getMasterInitializedTime`.
 - Aggiunta patch che migliora il tempo di avvio del cluster.

Problemi noti

- L'avvio del cluster non va a buon fine quando tutte le applicazioni sono installate e la dimensione del volume di root Amazon EBS di default non viene modificata. Per risolvere il problema, utilizza il comando `aws emr create-cluster` da AWS CLI e specifica un parametro `--ebs-root-volume-size` più grande.
- Per impostazione predefinita, Hive 2.3.0 è impostato su `hive.compute.query.using.stats=true`. Questo fa sì che le query individuino i dati dalle statistiche esistenti anziché direttamente dai dati, il che potrebbe creare confusione. Ad esempio,

se hai una tabella con `hive.compute.query.using.stats=true` e carichi nuovi file sulla tabella `LOCATION`, se esegui una query `SELECT COUNT(*)` sulla tabella ti restituisce il conto delle statistiche e non tiene in considerazione le righe aggiunte.

Una soluzione consiste nell'usare il comando `ANALYZE TABLE` per raccogliere nuove statistiche o impostare `hive.compute.query.using.stats=false`. Per ulteriori informazioni, consulta [Statistiche in Hive](#) nella documentazione di Apache Hive.

- Spark: quando si utilizza Spark, c'è un problema di perdita del gestore di file con il daemon `appusher`, che può apparire per un processo Spark a esecuzione prolungata dopo diverse ore o giorni. Per risolvere il problema, connettiti al nodo master e digita `sudo /etc/init.d/appusher stop`. Questa azione arresta l'appusher daemon, che Amazon EMR riavvierà automaticamente.
- Cronologia dell'applicazione
 - I dati cronologici per gli executor Spark terminati non sono disponibili.
 - La cronologia dell'applicazione non è disponibile per i cluster che utilizzano una configurazione di sicurezza per attivare la crittografia in-flight.

Rilascio 5.7.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.7.0. Le modifiche sono relative al rilascio di Amazon EMR 5.6.0.

Data del rilascio: 13 luglio 2017

Aggiornamenti

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Nuove funzionalità di

- Aggiunta la capacità di specificare un Amazon Linux AMI quando crei un cluster. Per ulteriori informazioni, consulta [Utilizzo di un'AMI personalizzata](#).

Modifiche, miglioramenti e problemi risolti

- HBase
 - Aggiunta la capacità di configurare la replica di lettura ai cluster di HBase. Consulta [Utilizzo della replica di lettura al cluster](#).
 - Numerosi miglioramenti e correzioni dei bug.
- Presto: aggiunta la possibilità di configurare `node.properties`.
- YARN: aggiunta la possibilità di configurare `container-log4j.properties`.
- Sqoop: backport per [SQOOP-2880](#), che introduce un argomento che ti permette di impostare la directory temporanea di Sqoop.

Rilascio 5.6.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.6.0. Le modifiche sono relative al rilascio di Amazon EMR 5.5.0.

Data del rilascio: 5 giugno 2017

Aggiornamenti

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. Questa è la prima versione di Mahout che supporta Spark 2.x in Amazon EMR versione 5.0 e successive.
- Spark 2.1.1

Modifiche, miglioramenti e problemi risolti

- Presto
 - Aggiunta la capacità di attivare la comunicazione sicura SSL/TLS tra i nodi Presto attivando la cifratura in transito tramite la configurazione di sicurezza. Per maggiori informazioni, consulta [Cifratura dei dati in transito](#).
 - Backport per [Presto 7661](#) che aggiunge l'opzione `VERBOSE` all'istruzione `EXPLAIN ANALYZE` per comunicare statistiche più dettagliate di basso livello relative a un piano di query.

Rilascio 5.5.3

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.5.3. Le modifiche sono relative alla versione 5.5.2.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Rilascio 5.5.2

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 5.5.2. Le modifiche sono relative alla versione 5.5.1.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Rilascio 5.5.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.5.1. Le modifiche sono relative al rilascio di Amazon EMR 5.5.0.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Rilascio 5.5.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.5.0. Le modifiche sono relative al rilascio di Amazon EMR 5.4.0.

Data del rilascio: 26 aprile 2017

Aggiornamenti

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Modifiche, miglioramenti e problemi risolti

- Spark
 - Backport per Spark Patch ([SPARK-20115](#)) [Fix DAGScheduler per ricalcolare tutti i blocchi shuffle perduti durante la non disponibilità del servizio shuffle esterno](#) alla versione 2.1.0 di Spark, che è inclusa in questo rilascio.
- Flink
 - Flink ora è integrato con Scala 2.11. Se utilizzi le API e le librerie Scala, ti consigliamo di utilizzare Scala 2.11 nei tuoi progetti.
 - Gestito un problema in cui i valori predefiniti di HADOOP_CONF_DIR e YARN_CONF_DIR non erano impostati correttamente, per cui `start-scala-shell.sh` non funzionava. Aggiunta inoltre la capacità di impostare tali valori usando `env.hadoop.conf.dir` e `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` o la classificazione della configurazione `flink-conf`.
 - Introdotta un nuovo comando specifico per EMR, `flink-scala-shell` come wrapper per `start-scala-shell.sh`. Ti consigliamo di usare questo comando anziché `start-scala-shell`. Il nuovo comando semplifica l'esecuzione. Ad esempio, `flink-scala-shell -n 2` avvia uno shell Flink Scala con un parallelismo dell'attività di 2.
 - Introdotta un nuovo comando specifico per EMR, `flink-yarn-session` come wrapper per `yarn-session.sh`. Ti consigliamo di usare questo comando anziché `yarn-session`. Il nuovo comando semplifica l'esecuzione. Ad esempio, `flink-yarn-session -d -n 2` avvia una sessione Flink di lunga durata in uno stato scollegato con due task manager.

- Risolto ([FLINK-6125 Commons httpclient non più schermato in Flink 1.2](#))
- Presto
 - Aggiunto supporto per l'autenticazione LDAP. L'utilizzo di LDAP con Presto on Amazon EMR richiede di abilitare l'accesso HTTPS per il coordinatore di Presto (`http-server.https.enabled=true` in `config.properties`). Per dettagli sulla configurazione, consulta [LDAP authentication \(Autenticazione LDAP\)](#) nella documentazione di Presto.
 - Aggiunta del supporto per SHOW GRANTS.
- AMI Linux di base di Amazon EMR
 - I rilasci di Amazon EMR sono ora basati su Amazon Linux 2017.03. Per ulteriori informazioni, consulta [Note di rilascio dell'AMI Amazon Linux 2017.03](#).
 - È stato rimosso Python 2.6 dall'immagine Linux di base Amazon EMR. Python 2.7 e 3.4 vengono installati per impostazione predefinita. Puoi installare Python 2.6 manualmente se necessario.

Rilascio 5.4.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.4.0. Le modifiche sono relative al rilascio di Amazon EMR 5.3.0.

Data di rilascio: 08 marzo 2017

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Flink 1.2.0
- Aggiornato a HBase 1.3.0
- Aggiornato a Phoenix 4.9.0

Note

Se effettui l'aggiornamento da una versione precedente di Amazon EMR ad Amazon EMR versione 5.4.0 o successive e utilizzi un'indicizzazione secondaria, effettua l'aggiornamento degli indici locali come descritto nella documentazione su [Apache Phoenix](#). Amazon EMR rimuove le configurazioni richieste dalla classificazione `hbase-site`, ma gli indici devono essere ripopolati. Gli aggiornamenti online e offline degli indici sono supportati. Gli aggiornamenti online sono quelli predefiniti, il che significa che gli indici

vengono ripopolati durante la fase di inizializzazione dai client Phoenix della versione 4.8.0 o più grande. Per specificare gli aggiornamenti offline, imposta la configurazione `phoenix.client.localIndexUpgrade` su falso nella classificazione `phoenix-site` e poi SSH sul nodo master per eseguire `psql [zookeeper] -1`.

- Aggiornato a Presto 0.166
- Aggiornato a Zeppelin 0.7.0

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio `emr-5.4.0`:

- Aggiunto supporto per le istanze `r4`. Consulta [Tipi di istanza Amazon EC2](#).

Rilascio 5.3.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.3.1. Le modifiche sono relative al rilascio di Amazon EMR 5.3.0.

Data di rilascio: 7 febbraio 2017

Modifiche minori alle patch Zeppelin di backport e aggiornamento dell'AMI predefinita per Amazon EMR.

Rilascio 5.3.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.3.0. Le modifiche sono relative al rilascio di Amazon EMR 5.2.1.

Data del rilascio: 26 gennaio 2017

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Hive 2.1.1
- Aggiornato a Hue 3.11.0
- Aggiornato a Spark 2.1.0

- Aggiornato a Oozie a 4.3.0
- Aggiornato a Flink 1.1.4

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-5.3.0:

- Aggiunta una patch a Hue che ti permette di utilizzare l'impostazione `interpreters_shown_on_wheel` per configurare che interpreti mostrare per primi sulla ruota di selezione Notebook, indipendentemente dall'ordine nel file `hue.ini`.
- Aggiunta la classificazione della configurazione `hive-parquet-logging`, che puoi utilizzare per configurare i valori nel file `parquet-logging.properties` di Hive.

Rilascio 5.2.2

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.2. Le modifiche sono relative al rilascio di Amazon EMR 5.2.1.

Data di rilascio: 2 maggio 2017

Problemi noti risolti dai rilasci precedenti

- Backport per [SPARK-194459](#), che risolve un problema in cui la lettura da una tabella ORC con colonne `char`/`varchar` restituisce un errore.

Rilascio 5.2.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.1. Le modifiche sono relative al rilascio di Amazon EMR 5.2.0.

Data del rilascio: 29 dicembre 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Presto 0.157.1. Per ulteriori informazioni, consulta [Presto release notes \(Note di rilascio di Presto\)](#) nella documentazione di Presto.

- Aggiornato a Zookeeper 3.4.9. Per ulteriori informazioni, consulta [ZooKeeper release notes \(Note di rilascio di ZooKeeper\)](#) nella documentazione di Apache ZooKeeper.

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-5.2.1:

- Aggiunto supporto per il tipo di istanza Amazon EC2 m4.16xlarge in Amazon EMR versione 4.8.3 e successive, escluse 5.0.0, 5.0.3 e 5.2.0.
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- La posizione dei percorsi di configurazione Flink e YARN è ora impostata per default in `/etc/default/flink` e non occorre impostare le variabili di ambiente `FLINK_CONF_DIR` e `HADOOP_CONF_DIR` quando si eseguono gli script dei driver `flink` o `yarn-session.sh` per avviare processi Flink.
- Aggiunto supporto per la classe `FlinkKinesisConsumer`.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in Hadoop in cui il thread `ReplicationMonitor` poteva bloccarsi per molto tempo a causa di un conflitto tra la replica e la cancellazione dello stesso file in un cluster di grandi dimensioni.
- Risolto un problema in cui `ControlledJob#toString` restituiva un errore con una null pointer exception (NPE) quando lo stato del processo non era aggiornato correttamente.

Rilascio 5.2.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.2.0. Le modifiche sono relative al rilascio di Amazon EMR 5.1.0.

Data del rilascio: 21 novembre 2016

Modifiche e miglioramenti

Le seguenti modifiche e miglioramenti sono disponibili in questo rilascio:

- Aggiunta la modalità di archiviazione Amazon S3 per HBase.
- Ti permette di specificare un percorso Amazon S3 per la rootdir HBase. Per ulteriori informazioni, consulta [HBase su Amazon S3](#).

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Spark 2.0.2

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con /mnt vincolato a 2 TB sui tipi di istanza solo EBS.
- Risolto un problema con il controller dell'istanza e i log logpusher che venivano indirizzati ai file .out corrispondenti anziché ai normali file .log log4j-configured, che ruotano ogni ora. I file .out non ruotano e questo determinerà il riempimento della partizione /emr. Questo problema interessa soltanto i tipi di istanza hardware virtual machine (HVM).

Rilascio 5.1.0

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.1.0. Le modifiche sono relative al rilascio di Amazon EMR 5.0.0.

Data del rilascio: 03 novembre 2016

Modifiche e miglioramenti

Le seguenti modifiche e miglioramenti sono disponibili in questo rilascio:

- Aggiunto supporto per Flink 1.1.3.
- Presto è stato aggiunto come opzione nella sezione notebook di Hue.

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a HBase 1.2.3

- Aggiornato a Zeppelin 0.6.2

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con le query Tez su Amazon S3 con i file ORC: la query non veniva eseguita correttamente come nelle precedenti versioni di Amazon EMR 4.x.

Rilascio 5.0.3

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 5.0.3. Le modifiche sono relative al rilascio di Amazon EMR 5.0.0.

Data del rilascio: 24 ottobre 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Hadoop 2.7.3
- Aggiornato a Presto 0.152.3, che include il supporto per l'interfaccia Web di Presto. Puoi accedere all'interfaccia Web di Presto sulla porta di utilizzo del coordinatore di Presto 8889. Per ulteriori informazioni sull'interfaccia Web di Presto, consulta [Web interface \(Interfaccia Web\)](#) nella documentazione di Presto.
- Aggiornato a Spark 2.0.1
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Rilascio 5.0.0

Data del rilascio: 27 luglio 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Hive 2.1
- Aggiornato a Presto 0.150

- Aggiornato a Spark 2.0
- Aggiornato a Hue 3.10.0
- Aggiornato a Pig 0.16.0
- Aggiornato a Tez 0.8.4
- Aggiornato a Zeppelin 0.6.1

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-5.0.0 o superiore:

- Amazon EMR supporta le ultime versioni open source di Hive (versione 2.1) e di Pig (versione 0.16.0). Se in passato hai utilizzato Hive o Pig su Amazon EMR, questo potrebbe interessare alcuni casi d'uso. Per ulteriori informazioni, consulta [Hive](#) e [Pig](#).
- Il motore di esecuzione di default per Hive e Pig ora è Tez. Per modificarlo, devi cambiare i valori appropriati nelle classificazioni della configurazione rispettivamente di `hive-site` e `pig-properties`.
- Una caratteristica di debugging della fase potenziata è stata aggiunta per permetterti di vedere la causa sottostante degli errori della fase se il servizio è in grado di determinare la causa. Per ulteriori informazioni, consulta [Debug migliorato delle fasi](#) nella Guida alla gestione di Amazon EMR.
- Le applicazioni che in precedenza terminavano con "-Sandbox" non avranno più quel suffisso. Questo potrebbe spezzare la tua automazione, ad esempio, se stai utilizzando gli script per lanciare i cluster con queste applicazioni. La seguente tabella mostra i nomi dell'applicazione in Amazon EMR 4.7.2 rispetto ad Amazon EMR 5.0.0.

Modifiche al nome dell'applicazione

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin

Amazon EMR 4.7.2	Amazon EMR 5.0.0
ZooKeeper-Sandbox	ZooKeeper

- Spark è compilato per Scala 2.11.
- Java 8 è il JVM di default. Tutte le applicazioni vengono eseguite utilizzando il runtime di Java 8. Non ci sono modifiche al target del bytecode dell'applicazione. La maggior parte delle applicazioni continua a scegliere come target Java 7.
- Zeppelin ora include le caratteristiche di autenticazione. Per ulteriori informazioni, consulta [Zeppelin](#).
- Aggiunto supporto per le configurazioni di sicurezza che ti permettono di creare e applicare le opzioni di cifratura più facilmente. Per ulteriori informazioni, consulta [Crittografia dei dati](#).

Rilascio 4.9.5

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 4.9.5. Le modifiche sono relative alla versione 4.9.4.

Data del rilascio iniziale: 29 agosto 2018

Modifiche, miglioramenti e problemi risolti

- HBase
 - Questo rilascio risolve una potenziale vulnerabilità di sicurezza.

Rilascio 4.9.4

Le seguenti note di rilascio includono informazioni relative al rilascio di Amazon EMR 4.9.4. Le modifiche sono relative alla versione 4.9.3.

Data del rilascio iniziale: 29 marzo 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere potenziali vulnerabilità.

Rilascio 4.9.3

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.3. Le modifiche sono relative al rilascio di Amazon EMR 4.9.2.

Data del rilascio iniziale: 22 gennaio 2018

Modifiche, miglioramenti e problemi risolti

- Aggiornato il kernel Amazon Linux dell'AMI default Amazon Linux per Amazon EMR per risolvere le vulnerabilità associate all'esecuzione speculativa (CVE-2017-5715, CVE-2017-5753 e CVE-2017-5754). Per ulteriori informazioni, consulta <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Rilascio 4.9.2

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.2. Le modifiche sono relative al rilascio di Amazon EMR 4.9.1.

Data del rilascio: 13 luglio 2017

Modifiche minori, correzioni dei bug e miglioramenti sono stati effettuati in questo rilascio.

Rilascio 4.9.1

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.9.1. Le modifiche sono relative al rilascio di Amazon EMR 4.8.4.

Data del rilascio: 10 aprile 2017

Problemi noti risolti dai rilasci precedenti

- Backport di [HIVE-9976](#) e [HIVE-10106](#)
- Risolto un problema in YARN in cui un numero elevato di nodi (superiore a 2000) e di container (superiore a 5000) causa un errore di esaurimento della memoria, ad esempio: "Exception in thread 'main' java.lang.OutOfMemoryError".

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-4.9.1:

- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2017.03. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- È stato rimosso Python 2.6 dall'immagine Linux di base Amazon EMR. Puoi installare Python 2.6 manualmente se necessario.

Rilascio 4.8.4

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.4. Le modifiche sono relative al rilascio di Amazon EMR 4.8.3.

Data di rilascio: 7 febbraio 2017

Modifiche minori, correzioni dei bug e miglioramenti sono stati effettuati in questo rilascio.

Rilascio 4.8.3

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.3. Le modifiche sono relative al rilascio di Amazon EMR 4.8.2.

Data del rilascio: 29 dicembre 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Presto 0.157.1. Per ulteriori informazioni, consulta [Presto release notes \(Note di rilascio di Presto\)](#) nella documentazione di Presto.
- Aggiornato a Spark 1.6.3. Per ulteriori informazioni, consulta [Spark release notes \(Note di rilascio di Spark\)](#) nella documentazione di Apache Spark.
- Aggiornato a Zookeeper 3.4.9. Per ulteriori informazioni, consulta [ZooKeeper release notes \(Note di rilascio di ZooKeeper\)](#) nella documentazione di Apache ZooKeeper.

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-4.8.3:

- Aggiunto supporto per il tipo di istanza Amazon EC2 m4.16xlarge in Amazon EMR versione 4.8.3 e successive, escluse 5.0.0, 5.0.3 e 5.2.0.
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in Hadoop in cui il thread ReplicationMonitor poteva bloccarsi per molto tempo a causa di un conflitto tra la replica e la cancellazione dello stesso file in un cluster di grandi dimensioni.
- Risolto un problema in cui ControlledJob#toString restituiva un errore con una null pointer exception (NPE) quando lo stato del processo non era aggiornato correttamente.

Rilascio 4.8.2

Le seguenti note di rilascio includono informazioni per il rilascio di Amazon EMR 4.8.2. Le modifiche sono relative al rilascio di Amazon EMR 4.8.0.

Data del rilascio: 24 ottobre 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a Hadoop 2.7.3
- Aggiornato a Presto 0.152.3, che include il supporto per l'interfaccia Web di Presto. Puoi accedere all'interfaccia Web di Presto sulla porta di utilizzo del coordinatore di Presto 8889. Per ulteriori informazioni sull'interfaccia Web di Presto, consulta [Web interface \(Interfaccia Web\)](#) nella documentazione di Presto.
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.09. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Rilascio 4.8.0

Data del rilascio: 7 settembre 2016

Aggiornamenti

I seguenti aggiornamenti sono disponibili in questo rilascio:

- Aggiornato a HBase 1.2.2
- Aggiornato a Presto-Sandbox 0.151
- Aggiornato a Tez 0.8.4
- Aggiornato a Zeppelin-Sandbox 0.6.1

Modifiche e miglioramenti

Di seguito sono riportate le modifiche apportate ai rilasci Amazon EMR per l'etichetta di rilascio emr-4.8.0:

- Risolto un problema con YARN dove ApplicationMaster cercava di pulire i container che non esistono più perché le loro istanze sono state terminate.
- Corrette le azioni hive-server2 URL per Hive2 negli esempi Oozie.
- Aggiunto supporto per i cataloghi Presto aggiuntivi.
- Backport per patch: [HIVE-8948](#), [HIVE-12679](#), [HIVE-13405](#), [PHOENIX-3116](#), [HADOOP-12689](#)
- Aggiunto supporto per le configurazioni di sicurezza che ti permettono di creare e applicare le opzioni di cifratura più facilmente. Per ulteriori informazioni, consulta [Crittografia dei dati](#).

Rilascio 4.7.2

Le seguenti note di rilascio includono informazioni su Amazon EMR 4.7.2.

Data del rilascio: 15 luglio 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiornato a Mahout 0.12.2

- Aggiornato a Presto 0.148
- Aggiornato a Spark 1.6.2
- Ora puoi creare un `AWSCredentialsProvider` da utilizzare con EMRFS utilizzando un URI come parametro. Per ulteriori informazioni, consulta [Crea un AWSCredentialsProvider per EMRFS](#).
- EMRFS ora consente agli utenti di configurare un endpoint DynamoDB personalizzato per i metadata Consistent View (Visualizzazione coerente) utilizzando la proprietà `fs.s3.consistent.dynamodb.endpoint` in `emrfs-site.xml`.
- Aggiunto uno script in `/usr/bin` chiamato `spark-example`, che esegue il wrapping di `/usr/lib/spark/spark/bin/run-example` in modo che tu possa eseguire gli esempi direttamente. Ad esempio, per eseguire un esempio SparkPi disponibile con la distribuzione Spark, puoi eseguire `spark-example SparkPi 100` dalla riga di comando o utilizzando `command-runner.jar` come fase nella API.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui Oozie aveva `spark-assembly.jar` in una posizione non corretta quando Spark era installato, il che portava a un errore durante il lancio delle applicazioni Spark con Oozie.
- Risolto un problema con il logging Spark Log4j-based nei container YARN.

Rilascio 4.7.1

Data del rilascio: 10 giugno 2016

Problemi noti risolti dai rilasci precedenti

- Risolto un problema che prolungava i tempi di avvio dei cluster lanciati in un VPC con sottoreti private. Il bug interessava soltanto i cluster lanciati con il rilascio di Amazon EMR 4.7.0.
- Risolto un problema che gestiva in modo non corretto l'elenco dei file in Amazon EMR per i cluster lanciati con il rilascio di Amazon EMR 4.7.0.

Rilascio 4.7.0

Important

Amazon EMR 4.7.0 è una versione obsoleta. Utilizza invece Amazon EMR 4.7.1 o versioni successive.

Data del rilascio: 2 giugno 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiunto Apache Phoenix 4.7.0
- Aggiunto Apache Tez 0.8.3
- Aggiornato a HBase 1.2.1
- Aggiornato a Mahout 0.12.0
- Aggiornato a Presto 0.147
- Aggiornato AWS SDK for Java a 1.10.75
- Il flag finale è stato rimosso dalla proprietà `mapreduce.cluster.local.dir` in `mapred-site.xml` per permettere agli utenti di eseguire Pig in modalità locale.

Driver JDBC Amazon Redshift disponibili sul cluster

I driver JDBC Amazon Redshift sono ora inclusi in `/usr/share/aws/redshift/jdbc./usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` è il driver JDBC Amazon Redshift compatibile con 4.1 e `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` è il driver JDBC Amazon Redshift compatibile con 4.0. Per ulteriori informazioni, consultare [Configurazione di una connessione JDBC](#) nella Guida alla gestione di Amazon Redshift.

Java 8

A eccezione di Presto, OpenJDK 1.7 è il JDK di default utilizzato per tutte le applicazioni. Tuttavia, sia OpenJDK 1.7 sia OpenJDK 1.8 sono installati. Per informazioni su come impostare `JAVA_HOME` per le applicazioni, consulta [Configurazione delle applicazioni per utilizzare Java 8](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con il kernel che influiva in modo significativo sulla performance dei volumi Throughput Optimized HDD (st1) EBS per Amazon EMR in emr-4.6.0.
- Risolto un problema in cui un cluster restituiva un errore se una zona di cifratura HDFS veniva specificata senza scegliere Hadoop come applicazione.
- Modificata la policy di scrittura HDFS di default da RoundRobin a AvailableSpaceVolumeChoosingPolicy. Alcuni volumi non venivano utilizzati correttamente con la configurazione RoundRobin, il che portava i nodi principali a restituire un errore e alla creazione di HDFS non affidabili.
- Risolto un problema con la CLI EMRFS che causava un'eccezione durante la creazione della tabella dei metadati DynamoDB di default per viste uniformi.
- Risolto un problema di deadlock in EMRFS che accadeva eventualmente durante le operazioni multipartite di rinomina e copia.
- Risolto un problema con EMRFS che faceva sì che la dimensione di default di CopyPart fosse 5 MB. Il valore di default è ora di 128 MB.
- Risolto un problema con la configurazione upstart di Zeppelin che eventualmente ti impediva di poter arrestare il servizio.
- Risolto un problema con Spark e Zeppelin, che ti impediva di utilizzare lo schema URI s3a:// perché /usr/lib/hadoop/hadoop-aws.jar non era caricato correttamente nel rispettivo classpath.
- Backport per [HUE-2484](#).
- Backport per [commit](#) da Hue 3.9.0 (nessun JIRA esistente) per risolvere un problema con il campione del browser HBase.
- Backport per [HIVE-9073](#).

Rilascio 4.6.0

Data del rilascio: 21 aprile 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiunto HBase 1.2.0

- Aggiunto Zookeeper-Sandbox 3.4.8
- Aggiornato a Presto-Sandbox 0.143
- I rilasci di Amazon EMR sono ora basati su Amazon Linux 2016.03.0. Per ulteriori informazioni, consulta <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.

Problema che riguarda i tipi di volume Throughput Optimized HDD (st1) EBS

Un problema nel kernel Linux versioni 4.2 e superiori influisce in modo significativo sulla performance dei volumi Throughput Optimized HDD (st1) EBS per EMR. Questo rilascio (emr-4.6.0) utilizza kernel versione 4.4.5 per cui è interessato dal problema. Per questo motivo ti consigliamo di non usare emr-4.6.0 se desideri usare i volumi st1 EBS. Puoi utilizzare i rilasci emr-4.5.0 o Amazon EMR precedenti con st1 senza che si crei alcun problema. Inoltre, forniamo la correzione per i rilasci futuri.

Valori di default di Python

Python 3.4 ora è installato di default, ma Python 2.7 rimane la versione di default del sistema. Puoi configurare Python 3.4 come la versione di default del sistema utilizzando un'operazione di bootstrap; puoi anche usare la configurazione API per impostare l'export `PYSPARK_PYTHON` su `/usr/bin/python3.4` nella classificazione `spark-env` in modo che interessi la versione di Python utilizzata da PySpark.

Java 8

A eccezione di Presto, OpenJDK 1.7 è il JDK di default utilizzato per tutte le applicazioni. Tuttavia, sia OpenJDK 1.7 sia OpenJDK 1.8 sono installati. Per informazioni su come impostare `JAVA_HOME` per le applicazioni, consulta [Configurazione delle applicazioni per utilizzare Java 8](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui il provisioning dell'applicazione poteva restituire un errore in modo casuale a causa della password generata.
- In precedenza, `mysqld` era installato su tutti i nodi. Ora, è installato soltanto sull'istanza master e solo se l'applicazione scelta include `mysql-server` come componente. Attualmente, le seguenti applicazioni includono il componente `mysql-server`: HCatalog, Hive, Hue, Presto-Sandbox e Sqoop-Sandbox.
- Modificato `yarn.scheduler.maximum-allocation-vcores` a 80 rispetto al valore di default di 32. Questa scelta risolve un errore introdotto in emr-4.4.0 che si verificava principalmente

con Spark durante l'uso dell'opzione `maximizeResourceAllocation` in un cluster il cui tipo di istanza core era uno dei pochi tipi di istanza di grandi dimensioni che hanno i vcore YARN impostati su un valore superiore a 32; tra questi `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` o `m4.10xlarge` erano interessati dal problema.

- `s3-dist-cp` ora utilizza EMRFS per tutte le nomine Amazon S3 e non viene più gestito su una directory HDFS temporanea.
- Risolto un problema con la gestione dell'eccezione per il caricamento in più parti della crittografia lato client.
- Aggiunta un'opzione per permettere agli utenti di modificare la classe di archiviazione Amazon S3. Per default, l'impostazione è `STANDARD`. L'impostazione della classificazione della configurazione `emrfs-site` è `fs.s3.storageClass` e i valori possibili sono `STANDARD`, `STANDARD_IA` e `REDUCED_REDUNDANCY`. Per ulteriori informazioni sulle classi di archiviazione, consulta [Classi di archiviazione](#) nella Guida per l'utente di Amazon Simple Storage Service.

Rilascio 4.5.0

Data del rilascio: 4 aprile 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiornato a Spark 1.6.1
- Aggiornato a Hadoop 2.7.2
- Aggiornato a Presto 0.140
- Aggiunto supporto AWS KMS per la crittografia lato server Amazon S3.

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui i server MySQL e Apache non si riavviavano dopo il riavvio di un nodo.
- Risolto un problema in cui `IMPORT` non funzionava correttamente con le tabelle non partizionate archiviate in Amazon S3
- Risolto un problema con Presto dove si richiedeva che la directory di gestione temporanea fosse `/mnt/tmp` anziché `/tmp` quando si scrive sulla tabelle Hive.

Rilascio 4.4.0

Data di rilascio: 14 marzo 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiunto HCatalog 1.0.0
- Aggiunto Sqoop-Sandbox 1.4.6
- Aggiornato a Presto 0.136
- Aggiornato a Zeppelin 0.5.6
- Aggiornato a Mahout 0.11.1
- Abilitato `dynamicResourceAllocation` per impostazione predefinita.
- Aggiunta una tabella di tutte le classificazioni della configurazione per il rilascio. Per ulteriori informazioni, consulta la tabella sulle classificazioni della configurazione in [Configurazione delle applicazioni](#).

Problemi noti risolti dai rilasci precedenti

- Risolto un problema in cui l'impostazione `maximizeResourceAllocation` non avrebbe riservato memoria sufficiente per i daemon YARN ApplicationMaster.
- Risolto un problema rilevato con un DNS personalizzato. Se le voci in `resolve.conf` precedono le voci personalizzate fornite, allora le voci personalizzate non sono risolvibili. Questo comportamento è stato influenzato dai cluster in un VPC dove il server di nomi VPC di default era inserito come prima voce in `resolve.conf`.
- Risolto un problema in cui il Python di default era spostato nella versione 2.7 e boto non era installato per quella versione.
- Risolto un problema in cui i container YARN e le applicazioni Spark generavano un file unico del round robin database (rrd) di Ganglia il che portava al collegamento del primo disco al riempimento dell'istanza. Per via di questa correzione, i parametri a livello del container YARN sono stati disattivati così come quelli a livello dell'applicazione Spark.
- Risolto un problema in log pusher che avrebbe cancellato tutte le cartelle vuote dei log. La conseguenza era che Hive CLI non era in grado di effettuare l'accesso perché log pusher cancellava la cartella vuota `user` sotto `/var/log/hive`.

- Risolto un problema che riguardava le importazioni di Hive, che interessava il partizionamento e produceva un errore durante l'importazione.
- Risolto un problema in cui EMRFS e s3-dist-cp non gestivano correttamente i nomi dei bucket che contengono punti.
- Modificato un comportamento in EMRFS che prevedeva che nei bucket con funzione Versioni multiple abilitata il file marker `_$folder$` non venisse creato in modo continuo, il che potrebbe portare a un miglioramento delle prestazioni per i bucket con funzione Versioni multiple abilitata.
- Modificato il comportamento in EMRFS in modo tale che non utilizzi i file di istruzione tranne nei casi in cui la crittografia lato client sia abilitata. Se vuoi cancellare i file di istruzione mentre utilizzi la crittografia lato client, puoi configurare la proprietà `emrfs-site.xml` property, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, su `true`.
- Modificata l'aggregazione dei log YARN per conservare i log nella destinazione dell'aggregazione per due giorni. La destinazione di default è lo storage HDFS del cluster. Se desideri modificare la durata, modifica il valore di `yarn.log-aggregation.retain-seconds` utilizzando la classificazione della configurazione `yarn-site` durante la creazione del cluster. Come sempre, puoi salvare i log dell'applicazione in Amazon S3 utilizzando il parametro `log-uri` quando crei il cluster.

Patch applicate

Le seguenti patch di progetti open source sono state incluse in questo rilascio:

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Rilascio 4.3.0

Data del rilascio: 19 gennaio 2016

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiornato a Hadoop 2.7.1
- Aggiornato a Spark 1.6.0

- Aggiornato Ganglia a 3.7.2
- Aggiornato Presto a 0.130

Amazon EMR ha effettuato alcune modifiche a `spark.dynamicAllocation.enabled` quando è impostato su `true` (vero); è `false` (falso) per impostazione predefinita. Quando è impostato su `true`, interessa i valori predefiniti definiti dall'impostazione `maximizeResourceAllocation`:

- Se `spark.dynamicAllocation.enabled` è impostato su `true`, `spark.executor.instances` non è impostato da `maximizeResourceAllocation`.
- L'impostazione `spark.driver.memory` ora è configurata in base ai tipi di istanza nel cluster in modo simile a come è impostato `spark.executors.memory`. Tuttavia, dato che l'applicazione del driver di Spark potrebbe essere eseguita o sull'istanza master o su una delle istanze core (ad esempio, nelle modalità client YARN e cluster, rispettivamente), l'impostazione `spark.driver.memory` è definita in base al tipo di istanza del tipo di istanza più piccola tra questi due gruppi di istanze.
- L'impostazione `spark.default.parallelism` è ora definita al doppio dei core CPU disponibili per i container YARN. Nei rilasci precedenti, era la metà del valore.
- I calcoli per il sovraccarico della memoria allocata ai processi YARN di Spark sono stati modificati per essere più precisi, generando un piccolo aumento della quantità di memoria disponibile per Spark (ovvero, `spark.executor.memory`).

Problemi noti risolti dai rilasci precedenti

- L'aggregazione dei log YARN ora è abilitata per impostazione predefinita.
- Risolto un problema in cui i log non venivano inviati al bucket di log Amazon S3 del cluster quando l'aggregazione dei log YARN era abilitata.
- Le dimensioni del container YARN ora sono di un minimo di 32 su tutti i tipi di nodo.
- Risolto un problema con Ganglia che provocava un I/O su disco eccessivo sul nodo master nei cluster di grandi dimensioni.
- Risolto un problema che evitava che i log delle applicazioni venissero trasmessi ad Amazon S3 quando un cluster veniva disattivato.
- Risolto un problema in EMRFS CLI che faceva sì che alcuni comandi restituissero errori.
- Risolto un problema con Zeppelin che evitava che le dipendenze venissero caricate nello `SparkContext` sottostante.

- Risolto un problema causato dall'emissione di un tentativo di ridimensionamento per aggiungere le istanze.
- Risolto un problema in Hive dove CREATE TABLE AS SELECT effettua troppe chiamate dell'elenco su Amazon S3.
- Risolto un problema in cui i cluster di grandi dimensioni non effettuano correttamente il provisioning quando sono installati Hue, Oozie e Ganglia.
- Risolto un problema in s3-dist-cp dove avrebbe restituito un codice di uscita diverso zero se avesse restituito un errore.

Patch applicate

Le seguenti patch di progetti open source sono state incluse in questo rilascio:

- [OOZIE-2402](#)
- [HIVE-12502](#)
- [HIVE-10631](#)
- [HIVE-12213](#)
- [HIVE-10559](#)
- [HIVE-12715](#)
- [HIVE-10685](#)

Rilascio 4.2.0

Data del rilascio: 18 novembre 2015

Funzionalità

Le seguenti caratteristiche sono disponibili in questo rilascio:

- Aggiunto supporto per Ganglia
- Aggiornato a Spark 1.5.2
- Aggiornato a Presto 0.125
- Aggiornato Oozie a 4.2.0
- Aggiornato Zeppelin a 0.5.5
- Aggiornato AWS SDK for Java a 1.10.27

Problemi noti risolti dai rilasci precedenti

- Risolto un problema con EMRFS CLI dove non utilizzava il nome della tabella dei metadata di default.
- Risolto un problema che avveniva durante l'uso delle tabelle supportate da ORC in Amazon S3.
- Risolto un problema che avveniva con una mancanza di corrispondenza della versione di Python nella configurazione di Spark.
- Risolto un problema quando lo stato di un nodo YARN non riusciva a effettuare il report a causa di problemi di DNS per i cluster in un VPC.
- Risolto un problema che si verificava quando YARN disattivava i nodi, generando applicazioni bloccate o impedendo di programmare nuove applicazioni.
- Risolto un problema che si verificava quando i cluster terminavano con lo stato `TIMED_OUT_STARTING`.
- Risolto un problema che si verificava quando si include la dipendenza di EMRFS Scala in altre compilazioni. La dipendenza da Scala è stata rimossa.

Configurazione delle applicazioni

Per sovrascrivere le configurazioni predefinite per un'applicazione, puoi fornire un oggetto di configurazione. Puoi utilizzare una sintassi abbreviata per fornire la configurazione o puoi fare riferimento all'oggetto di configurazione in un file JSON. Gli oggetti di configurazione sono composti da una classificazione, proprietà e configurazioni nidificate opzionali. Le proprietà corrispondono alle impostazioni dell'applicazione che intendi utilizzare. Puoi specificare diverse classificazioni per più applicazioni in un singolo oggetto JSON.

Warning

Le operazioni API Describe e List di Amazon EMR emettono impostazioni personalizzate e configurabili, utilizzate come parte dei flussi di lavoro EMR di Amazon, come testo non formattato. Per fornire informazioni sensibili in queste impostazioni, ad esempio password, consulta [Come memorizzare dati di configurazione sensibili in AWS Secrets Manager](#).

Le classificazioni di configurazione disponibili variano a seconda della versione di rilascio di Amazon EMR. Per un elenco delle classificazioni di configurazione supportate in una versione di rilascio specifica, consulta la pagina della versione di rilascio in [Informazioni sui rilasci di Amazon EMR](#).

Di seguito è riportato un file JSON di esempio per un elenco di configurazioni:

```
[
  {
    "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"
    }
  }
]
```

Una classificazione di configurazione spesso esegue la mappatura a un file di configurazione specifico per l'applicazione. Ad esempio, la classificazione `hive-site` esegue la mappatura alle impostazioni nel file di configurazione `hive-site.xml` per Hive. Un'eccezione a questo esempio è l'operazione di bootstrap `configure-daemons`, non più supportata, che viene utilizzata per impostare parametri di ambiente come ad esempio `--namenode-heap-size`. Opzioni di questo genere sono inglobate nelle classificazioni `hadoop-env` e `yarn-env` con proprie classificazioni di esportazione nidificate. Se una classificazione termina in `env`, utilizza la classificazione secondaria di esportazione.

Un'altra eccezione è `s3get`, utilizzato per l'invio di un oggetto `EncryptionMaterialsProvider` del cliente in ogni nodo di un cluster per l'utilizzo nella crittografia lato client. A questo scopo è stata aggiunta un'opzione alla classificazione `emrfs-site`.

Di seguito è riportato un esempio della classificazione `hadoop-env`:

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Di seguito è riportato un esempio della classificazione `yarn-env`:

```
[
  {
    "Classification": "yarn-env",
```



```

    "Properties": {
    },
    "Configurations": [
    {
        "Classification": "export",
        "Properties": {
            "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdp:transport=dt_socket"
        },
        "Configurations": [
        ]
    }
    ]
}
]

```

Le seguenti impostazioni non appartengono a un file di configurazione, ma vengono utilizzate da Amazon EMR per configurare potenzialmente più impostazioni in modo automatico.

Impostazioni curate da Amazon EMR

Applicazione	Classificazione etichetta release	Proprietà valide	Quando usare
Spark	spark	maximizeResourceAllocation	Configurare gli esecutori per utilizzare le massime risorse di ogni nodo.

Argomenti

- [Configurazione delle applicazioni durante la creazione di un cluster](#)
- [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#)
- [Come memorizzare dati di configurazione sensibili in AWS Secrets Manager](#)
- [Configurazione delle applicazioni per l'uso di una macchina virtuale Java specifica](#)

Configurazione delle applicazioni durante la creazione di un cluster

Quando crei un cluster, puoi sovrascrivere la configurazione di default per le applicazioni utilizzando la console di Amazon EMR, AWS Command Line Interface (AWS CLI) o SDK AWS.

Per sovrascrivere la configurazione di default per un'applicazione, devi specificare i valori personalizzati in una classificazione di configurazione. Una classificazione di configurazione corrisponde a un file XML di configurazione per un'applicazione, ad esempio `hive-site.xml`.

Le classificazioni di configurazione variano a seconda della versione di rilascio di Amazon EMR. Per un elenco delle classificazioni di configurazione disponibili per ciascuna versione di rilascio, consulta la pagina dei dettagli del rilascio. Ad esempio, [Rilascio 6.4.0 di Amazon EMR](#).

Fornitura di una configurazione nella console quando si crea un cluster

Per specificare una configurazione, passa alla pagina Create cluster (Crea cluster) e scegli Edit software settings (Modifica impostazioni software). Puoi quindi immettere la configurazione direttamente utilizzando JSON o una sintassi abbreviata illustrata nel testo ombreggiato nella console. In caso contrario, puoi fornire un URI Amazon S3 per un file con un oggetto Configurations JSON.

Per specificare una configurazione per un gruppo di istanze, passa alla pagina Hardware Configuration (Configurazione hardware). Nella colonna Instance type (Tipo di istanza) della tabella Node type (Tipo di nodo), scegli di modificare le Configurations (Configurazioni) per applicazioni di ogni gruppo di istanze.

Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster

È possibile fornire una configurazione per il comando `create-cluster` specificando il percorso di un file JSON archiviato in locale o in Amazon S3. L'esempio seguente presuppone l'utilizzo dei ruoli predefiniti per Amazon EMR precedentemente creati. Se è necessario creare ruoli, esegui prima `aws emr create-default-roles`.

Se la configurazione è nella directory locale, puoi utilizzare il seguente comando di esempio:

```
aws emr create-cluster --use-default-roles --release-label emr-5.36.1 --applications
Name=Hive \
```

```
--instance-type m5.xlarge --instance-count 3 --configurations file:///./
configurations.json
```

Se la configurazione si trova in un percorso Simple Storage Service (Amazon S3), sarà necessario impostare la seguente soluzione alternativa prima di passare il percorso Simple Storage Service (Amazon S3) al comando `create-cluster`.

```
#!/bin/sh
# Assume the ConfigurationS3Path is not public, and its present in the same AWS account
as the EMR cluster
ConfigurationS3Path="s3://my-bucket/config.json"
# Get a presigned HTTP URL for the s3Path
ConfigurationURL=`aws s3 presign $ConfigurationS3Path --expires-in 300`
# Fetch the presigned URL, and minify the JSON so that it spans only a single line
Configurations=`curl $ConfigurationURL | jq -c .`
aws emr create-cluster --use-default-roles --release-label emr-5.34.0 --instance-type
m5.xlarge --instance-count 2 --applications Name=Hadoop Name=Spark --configurations
$Configurations
```

Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster

L'estratto del programma di seguito mostra come fornire una configurazione usando AWS SDK for Java:

```
Application hive = new Application().withName("Hive");

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration myHiveConfig = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Create cluster with ReleaseLabel")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(hive)
    .withConfigurations(myHiveConfig)
    .withServiceRole("EMR_DefaultRole")
```

```
.withJobFlowRole("EMR_EC2_DefaultRole")
.withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myEc2Key")
    .withInstanceCount(3)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m4.large")
    .withSlaveInstanceType("m4.large")
);
```

Riconfigurazione di un gruppo di istanze in un cluster in esecuzione

Con Amazon EMR versione 5.21.0 e successive, puoi riconfigurare le applicazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, puoi utilizzare la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS.

Quando aggiorni la configurazione di un'applicazione per un gruppo di istanze nella nuova console di Amazon EMR, la console tenta di unire la nuova configurazione con quella esistente per creare una nuova configurazione attiva. Nel caso insolito in cui Amazon EMR non riesca a unire la configurazione, ricevi un avviso da parte della console.

Dopo aver inviato una richiesta di riconfigurazione per un gruppo di istanze, Amazon EMR assegna un numero di versione alla nuova specifica di configurazione. Puoi monitorare il numero di versione di una configurazione o lo stato di un gruppo di istanze visualizzando gli eventi CloudWatch. Per ulteriori informazioni, consulta [Monitoraggio di CloudWatch Events](#).

Note

È possibile limitarsi a sovrascrivere, e non eliminare, le configurazioni cluster specificate durante la creazione del cluster. Se vi sono differenze tra la configurazione esistente e il file fornito, Amazon EMR avvia le operazioni di riconfigurazione, riavvia alcune applicazioni e reimposta tutte le configurazioni modificate manualmente, ad esempio le configurazioni modificate durante la connessione al cluster tramite SSH, ai valori di default per il gruppo di istanze specificato.

Considerazioni sulla riconfigurazione di un gruppo di istanze

Operazioni di riconfigurazione

Quando invii una richiesta di riconfigurazione utilizzando la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS, Amazon EMR controlla il file di configurazione esistente nel cluster. Se vi sono differenze tra la configurazione esistente e il file fornito, Amazon EMR avvia le operazioni di riconfigurazione, riavvia alcune applicazioni e reimposta tutte le configurazioni modificate manualmente, ad esempio le configurazioni modificate durante la connessione al cluster tramite SSH, ai valori di default per il gruppo di istanze specificato.

Note

Amazon EMR esegue alcune operazioni predefinite durante ogni riconfigurazione del gruppo di istanze. Queste azioni predefinite potrebbero entrare in conflitto con le personalizzazioni del cluster eseguite e causare errori di riconfigurazione. Per informazioni su come risolvere gli errori di riconfigurazione, consulta [Risoluzione dei problemi relativi alla configurazione del gruppo di istanze](#).

Amazon EMR avvia anche operazioni di riconfigurazione per le classificazioni di configurazione specificate nella richiesta. Per un elenco completo di queste operazioni, consulta la sezione [Classificazioni di configurazione per la versione di Amazon EMR in uso](#). Per esempio, [Classificazioni di configurazione 6.2.0](#).

Note

La Guida ai rilasci di Amazon EMR elenca solo le operazioni di riconfigurazione a partire dalle versioni di Amazon EMR 5.32.0 e 6.2.0.

Interruzione del servizio

Amazon EMR segue un processo in sequenza per riconfigurare le istanze nei gruppi di istanze attività e principali. Ogni volta, solo il 10% delle istanze in un gruppo di istanze vengono modificate e riavviate. Questo processo richiede più tempo per il completamento, ma riduce la possibilità di errori applicazione potenziali in un cluster in esecuzione.

Per eseguire processi YARN durante un riavvio di YARN, è possibile creare un cluster Amazon EMR con più nodi master o impostare `yarn.resourcemanager.recovery.enabled` su `true` nella classificazione di configurazione `yarn-site`. Per ulteriori informazioni sull'utilizzo di più nodi master, consulta [YARN ResourceManager a disponibilità elevata](#).

Convalida dell'applicazione

Amazon EMR verifica che ogni applicazione nel cluster sia in esecuzione dopo il processo di riavvio della riconfigurazione. Se un'applicazione non è disponibile, l'operazione di riconfigurazione complessiva ha esito negativo. Se un'operazione di riconfigurazione ha esito negativo, Amazon EMR ripristina i parametri di configurazione a quelli della versione di lavoro precedente.

Note

Per evitare errori di riconfigurazione, si consiglia di installare solo le applicazioni nel cluster che si prevede di utilizzare. Si consiglia inoltre di assicurarsi che tutte le applicazioni cluster siano integre e in esecuzione prima di inviare una richiesta di riconfigurazione.

Tipi di riconfigurazione

È possibile riconfigurare un gruppo di istanze attraverso uno dei due metodi descritti di seguito:

- **Sovrascrittura.** Il metodo di riconfigurazione predefinito e l'unico disponibile nei rilasci di Amazon EMR precedenti alla versione 5.35.0 e 6.6.0. Questo metodo di riconfigurazione sovrascrive in modo indiscriminato qualsiasi file sul cluster con il set di configurazione appena inviato. Inoltre, cancella qualsiasi modifica ai file di configurazione effettuata al di fuori dell'API di riconfigurazione.
- **Unione.** Metodo di riconfigurazione supportato per le release 5.35.0 e 6.6.0 o successive di Amazon EMR, tranne la console Amazon EMR, nella quale non è supportato da nessuna versione. Questo metodo di riconfigurazione unisce le configurazioni appena inviate con configurazioni già esistenti sul cluster. Questa opzione si limita ad aggiungere o modificare le nuove configurazioni che invii. Conserva le configurazioni esistenti.

 Note

Amazon EMR continua a sovrascrivere alcune configurazioni essenziali di Hadoop di cui necessita per garantire la corretta esecuzione del servizio.

Limitazioni

Quando riconfiguri un gruppo di istanze in un cluster in esecuzione, considera le limitazioni seguenti:

- Le applicazioni non YARN possono avere esito negativo durante il riavvio o causare problemi al cluster, soprattutto se le applicazioni non sono configurate correttamente. I cluster che si avvicinano al massimo utilizzo della memoria e della CPU potrebbero essere interessati da problemi dopo il processo di riavvio. Ciò vale soprattutto per il gruppo di istanze master.
- Non è possibile inviare una richiesta di riconfigurazione quando un gruppo di istanze viene ridimensionato. Se viene avviata una riconfigurazione durante il ridimensionamento di un gruppo di istanze, la riconfigurazione non può iniziare finché il ridimensionamento del gruppo di istanze non è terminato, e viceversa.
- Dopo la riconfigurazione di un gruppo di istanze, Amazon EMR riavvia le applicazioni per consentire alle nuove configurazioni di diventare effettive. Errori del processo o altri comportamenti dell'applicazione imprevisti possono verificarsi se le applicazioni sono in uso durante la riconfigurazione.
- Se una riconfigurazione per un gruppo di istanze ha esito negativo, Amazon EMR ripristina i parametri di configurazione a quelli della versione di lavoro precedente. Se il processo di ripristino ha esito negativo, è necessario inviare una nuova richiesta `ModifyInstanceGroup` per ripristinare il gruppo di istanze dallo stato `SUSPENDED`.
- Le richieste di riconfigurazione per le classificazioni di configurazione Phoenix sono supportate solo in Amazon EMR versione 5.23.0 e successive e non in Amazon EMR versione 5.21.0 o 5.22.0.
- Le richieste di riconfigurazione per le classificazioni di configurazione HBase sono supportate solo in Amazon EMR versione 5.30.0 e successive e non in Amazon EMR versioni da 5.23.0 a 5.29.0.
- Amazon EMR supporta le richieste di riconfigurazione delle applicazioni su un cluster Amazon EMR con più nodi primari solo nelle versioni 5.27.0 e successive di Amazon EMR.
- La riconfigurazione della classificazione `hdfs-encryption-zones` o di una qualsiasi classificazione di configurazione Hadoop KMS non è supportata su un cluster Amazon EMR con più nodi primari.

- Amazon EMR attualmente non supporta alcune richieste di riconfigurazione per il pianificatore di capacità che richiede il riavvio di YARN ResourceManager. Ad esempio, non è possibile rimuovere completamente una coda.

Riconfigurazione di un gruppo di istanze nella console

Note

La console Amazon EMR non supporta le riconfigurazioni di tipo Merge (Unione).

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/elasticmapreduce/>.
2. Nell'elenco di cluster sotto Name (Nome), scegli il cluster attivo che desideri riconfigurare.
3. Apri la pagina dei dettagli del cluster per il cluster e passa alla scheda Configurations (Configurazioni).
4. Nell'elenco a discesa Filter (Filtro), selezionare il gruppo di istanze da riconfigurare.
5. Nel menu a discesa Reconfigure (Riconfigura), scegli Edit in table (Modifica nella tabella) o Edit in JSON file (Modifica nel file JSON).
 - Edit in table (Modifica nella tabella): nella tabella di classificazioni di configurazione, modifica la proprietà e il valore per configurazioni esistenti o scegli Add configuration (Aggiungi configurazione) per specificare classificazioni di configurazione aggiuntive.
 - Edit in JSON file (Modifica nel file JSON): è possibile immettere la configurazione direttamente in formato JSON o utilizzare una sintassi abbreviata illustrata nel testo oscurato. In caso contrario, fornisci un URI Amazon S3 per un file con un oggetto Configurations JSON.

Note

La colonna Source (Origine) nella tabella di classificazioni di configurazione indica se la configurazione viene specificata al momento della creazione di un cluster o quando vengono specificate configurazioni aggiuntive per questo gruppo di istanze. È possibile modificare le configurazioni per un gruppo di istanze da entrambe le origini. Non è possibile eliminare le configurazioni cluster iniziali, ma è possibile sovrascriverle per un gruppo di istanze.

È anche possibile aggiungere o modificare classificazioni di configurazione nidificate direttamente nella tabella. Ad esempio, per fornire una classificazione secondaria `export` aggiuntiva di `hadoop-env`, aggiungere una classificazione di configurazione `hadoop.export` nella tabella. Quindi, fornire proprietà e valore specifici per questa classificazione.

6. (Facoltativo) Selezionare Apply this configuration to all active instance groups (Applica questa configurazione a tutti i gruppi di istanze attivi).
7. Salvare le modifiche.

Riconfigurazione di un gruppo di istanze utilizzando la CLI

Utilizza il comando `modify-instance-groups` per specificare una nuova configurazione per un gruppo di istanze in un cluster in esecuzione.

Note

Nei seguenti esempi, sostituisci `<j-2AL4XXXXXX5T9>` con l'ID cluster e sostituisci `<ig-1xxxxxxx9>` con l'ID del gruppo di istanze.

Example – Sostituzione di una configurazione per un gruppo di istanze

L'esempio seguente fa riferimento a un file JSON di configurazione denominato `instanceGroups.json` per modificare la proprietà del controllo dello stato del disco YARN NodeManager per un gruppo di istanze.

1. Prepara la classificazione della configurazione e salvala come `instanceGroups.json` nella stessa directory in cui esegui il comando.

```
[
  {
    "InstanceId": "<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. Esegui il comando seguente.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Aggiunta di una configurazione a un gruppo di istanze

Se desideri aggiungere una configurazione a un gruppo di istanze, devi includere le configurazioni specificate in precedenza per il gruppo di istanze nella nuova richiesta `ModifyInstanceGroup`. In caso contrario, le configurazioni specificate in precedenza vengono rimosse.

Nell'esempio seguente viene aggiunta una proprietà per il controllo della memoria virtuale di YARN NodeManager. La configurazione include anche i valori specificati in precedenza per il controllo dello stato del disco YARN NodeManager in modo che i valori non vengano sovrascritti.

1. Prepara i seguenti contenuti in `instanceGroups.json` e salvali nella stessa directory in cui esegui il comando:

```

[
  {
    "InstanceId": "<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. Esegui il comando seguente.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example - Aggiunta di una configurazione a un gruppo di istanze con una riconfigurazione di tipo Unione

Se intendi utilizzare il metodo di riconfigurazione predefinito Overwrite (Sovrascrittura) per aggiungere una configurazione, devi includere tutte le configurazioni specificate in precedenza per tale gruppo di istanze nella nuova richiesta `ModifyInstanceGroup`. Altrimenti, Overwrite (Sovrascrittura) rimuove le configurazioni precedentemente specificate. Questa operazione non è necessaria per la riconfigurazione Merge (Unione). Piuttosto devi accertarti che la tua richiesta includa solo le nuove configurazioni.

Nell'esempio seguente viene aggiunta una proprietà per il controllo della memoria virtuale di YARN NodeManager. Poiché si tratta di una riconfigurazione del tipo Merge (Unione), non sovrascrive i valori specificati in precedenza per il controllo dell'integrità del disco YARN NodeManager.

1. Prepara i seguenti contenuti in `instanceGroups.json` e salvali nella stessa directory in cui esegui il comando:

```

[
  {"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. Esegui il comando seguente.

```

aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json

```

Example – Eliminazione di una configurazione per un gruppo di istanze

Per eliminare una configurazione per un gruppo di istanze, invia una nuova richiesta di riconfigurazione che esclude la configurazione precedente.

Note

Puoi solo sostituire la configurazione del cluster iniziale. Non puoi eliminarla.

Ad esempio, per eliminare la configurazione per il controllo dello stato del disco YARN NodeManager specificato nell'esempio precedente, invia una nuova `instanceGroups.json` con i seguenti contenuti.

```

[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        }
      },
      "Configurations": []
    ]
  }
]

```

Note

Per eliminare tutte le configurazioni nell'ultima richiesta di riconfigurazione, invia una richiesta di riconfigurazione con un array di configurazioni vuoto. Ad esempio,

```
[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]
```

Example – Riconfigurazione e ridimensionamento di un gruppo di istanze in un'unica richiesta

Il JSON di esempio seguente illustra come riconfigurare e ridimensionare un gruppo di istanze nella stessa richiesta:

```
[
  {
    "InstanceId": "<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"
    }
}
]
}
]

```

Riconfigurazione di un gruppo di istanze utilizzando Java SDK

Note

Nei seguenti esempi, sostituisci `<j-2AL4XXXXXX5T9>` con l'ID cluster e sostituisci `<ig-1xxxxxxx9>` con l'ID del gruppo di istanze.

Il frammento di codice seguente fornisce una nuova configurazione per un gruppo di istanze utilizzando 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);
```

Il frammento di codice seguente elimina una configurazione specificata in precedenza per un gruppo di istanze specificando un array di configurazioni vuoto.

```
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);
```

Risoluzione dei problemi relativi alla configurazione del gruppo di istanze

Se il processo di riconfigurazione per un gruppo di istanze ha esito negativo, Amazon EMR ripristina la riconfigurazione e registra un messaggio di errore utilizzando un evento Amazon CloudWatch. L'evento fornisce un breve riepilogo dell'errore di riconfigurazione. Elenca le istanze per le quali la riconfigurazione non è riuscita e i messaggi di errore corrispondenti. Di seguito è riportato un esempio di messaggio di errore.

```
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".
```

Per raccogliere ulteriori dati su un errore di riconfigurazione, è possibile controllare i log di provisioning dei nodi. Ciò risulta particolarmente utile quando si riceve un messaggio come il seguente.

```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were
applied."
```

On the node

Accesso ai registri di provisioning dei nodi mediante connessione a un nodo

1. Utilizza SSH per connetterti al nodo sul quale la riconfigurazione non è riuscita. Per istruzioni, consulta [Connessione all'istanza Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
2. Passa alla directory seguente, che contiene i file di log di provisioning dei nodi.

```
/mnt/var/log/provision-node/
```

3. Apri la sottodirectory `reports` e cerca il report di provisioning dei nodi per la riconfigurazione. La directory `reports` organizza i log in base al numero di versione di riconfigurazione, all'identificatore univoco universale (UUID), all'indirizzo IP dell'istanza Amazon EC2 e al timestamp. Ogni report è un file YAML compresso che contiene informazioni dettagliate sul processo di riconfigurazione.

Di seguito è riportato un esempio di nome del file di report e relativo percorso.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxx/ip-10-73-xxx-xxx.ec2.internal/202104061715.yaml.gz
```

4. Puoi esaminare un report utilizzando un visualizzatore di file quale `zless`, come nel seguente esempio.

```
zless 202104061715.yaml.gz
```

Amazon S3

Accesso ai log di provisioning dei nodi tramite Amazon S3

1. Accedi alla AWS Management Console e apri la console Amazon S3 all'indirizzo <https://console.aws.amazon.com/s3/>.
2. Apri il bucket Amazon S3 specificato quando il cluster è stato configurato per archiviare i file di log.
3. Passa alla cartella seguente, che contiene i file di log di provisioning dei nodi:


```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-
node/
```

4. Apri la cartella `reports` e cerca il report di provisioning dei nodi per la riconfigurazione. La cartella `reports` organizza i log in base al numero di versione di riconfigurazione, all'identificatore univoco universale (UUID), all'indirizzo IP dell'istanza Amazon EC2 e al timestamp. Ogni report è un file YAML compresso che contiene informazioni dettagliate sul processo di riconfigurazione.

Di seguito è riportato un esempio di nome del file di report e relativo percorso.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

5. Per visualizzare un file di log, puoi scaricarlo da Amazon S3 sul tuo computer come file di testo. Per istruzioni, consulta [Download di un oggetto](#).

Ogni file di log contiene un report di provisioning dettagliato per la riconfigurazione associata. Per trovare informazioni sui messaggi di errore, cerca il livello di log `err` di un report. Il formato del report dipende dalla versione di Amazon EMR sul cluster.

L'esempio seguente mostra informazioni sugli errori per le versioni di Amazon EMR precedenti alla 5.32.0 e 6.2.0.

```
- !ruby/object:Puppet::Util::Log
  level: !ruby/sym err
  tags:
    - err
  message: "Example detailed error message."
  source: Puppet
  time: 2021-01-01 00:00:00.000000 +00:00
```

Le versioni di Amazon EMR 5.32.0, 6.2.0 e successive utilizzano invece il seguente formato.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
    - err
```

```
time: '2021-01-01 00:00:00.000000 +00:00'  
file:  
line:
```

Come memorizzare dati di configurazione sensibili in AWS Secrets Manager

Amazon EMR descrive ed elenca operazioni API che emettono dati di configurazione personalizzati (ad esempio `DescribeCluster` e `ListInstanceGroups`) in testo normale. Amazon EMR si integra con AWS Secrets Manager in modo che puoi memorizzare i tuoi dati in Secrets Manager e utilizzare l'ARN di un segreto nelle tue configurazioni. In tal modo, non trasmetti dati di configurazione sensibili ad Amazon EMR in testo normale e non li esponi ad API esterne. Se indichi che una coppia chiave-valore contiene un ARN di un segreto archiviato in Secrets Manager, Amazon EMR recupera questo segreto quando invia i dati di configurazione al cluster. Amazon EMR non invia l'annotazione quando utilizza API esterne per visualizzare la configurazione.

Creazione di un segreto

Per creare un segreto, segui la procedura in [Creazione di un segreto AWS Secrets Manager](#) nella Guida per l'utente di AWS Secrets Manager. Nel passaggio 3 devi scegliere il campo Plaintext (Testo normale) per inserire il tuo valore sensibile.

Tieni presente che mentre Secrets Manager consente a un segreto di contenere fino a 65536 byte, Amazon EMR limita a 1024 caratteri la lunghezza combinata della chiave della proprietà (esclusa l'annotazione) e del valore del segreto recuperato.

Come concedere ad Amazon EMR l'accesso per recuperare il segreto

Amazon EMR utilizza un ruolo di servizio IAM per il provisioning e la gestione automatica dei cluster. Il ruolo di servizio di Amazon EMR definisce le azioni consentite ad Amazon EMR quando effettua il provisioning delle risorse ed esegue attività a livello di servizio che non vengono eseguite nel contesto di un'istanza Amazon EC2 in esecuzione in un cluster. Per ulteriori informazioni sui ruoli di servizio, consulta [Ruolo di servizio per Amazon EMR \(ruolo EMR\)](#) e [Personalizzazione dei ruoli IAM](#).

Per consentire ad Amazon EMR di recuperare il valore del segreto da Secrets Manager, aggiungi la seguente dichiarazione di policy al tuo ruolo Amazon EMR quando avvii il cluster.

```
{
```

```

    "Sid": "AllowSecretsRetrieval",
    "Effect": "Allow",
    "Action": "secretsmanager:GetSecretValue",
    "Resource": [
      "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
    ]
  }

```

Se crei il segreto con una chiave AWS KMS key gestita dal cliente, per la chiave che usi devi aggiungere anche l'autorizzazione `kms:Decrypt` al ruolo Amazon EMR. Per ulteriori informazioni, consulta [Autenticazione e controllo degli accessi AWS Secrets Manager](#) nella Guida per l'utente di AWS Secrets Manager.

Come utilizzare il segreto in una classificazione di configurazione

Puoi aggiungere l'annotazione `EMR.secret@` a qualunque proprietà di configurazione per indicare che la coppia chiave-valore contiene un ARN di un segreto memorizzato in Secrets Manager.

Nell'esempio seguente viene mostrato come fornire l'ARN di un segreto in una classificazione di configurazione:

```

{
  "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>"
  }
}

```

Quando crei un cluster e invii la configurazione annotata, Amazon EMR convalida le proprietà di configurazione. Se la configurazione è valida, Amazon EMR elimina l'annotazione dalla configurazione e recupera il segreto da Secrets Manager per creare la configurazione effettiva prima di applicarla al cluster:

```

{
  "Classification": "core-site",
  "Properties": {
    "presto.s3.access-key": "<sensitive-access-key>",
    "presto.s3.secret-key": "<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}

```

```
}  
}
```

Quando invochi un'azione come `DescribeCluster`, Amazon EMR restituisce la configurazione dell'applicazione corrente sul cluster. Se una proprietà di configurazione dell'applicazione è contrassegnata come contenente l'ARN di un segreto, la configurazione dell'applicazione restituita dalla chiamata `DescribeCluster` contiene l'ARN e non il valore del segreto. Ciò garantisce che il valore del segreto sia visibile solo sul cluster:

```
{  
  "Classification": "core-site",  
  "Properties": {  
    "presto.s3.access-key": "<ensitive-access-key>",  
    "presto.s3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"  
  }  
}
```

Come aggiornare il valore del segreto

Amazon EMR recupera il valore del segreto da una configurazione annotata ogni volta che il gruppo di istanze collegato viene avviato, riconfigurato o ridimensionato. Puoi utilizzare Secrets Manager per modificare il valore di un segreto utilizzato nella configurazione di un cluster in esecuzione. In tal caso, puoi inviare una richiesta di riconfigurazione a ogni gruppo di istanze che devono ricevere il valore aggiornato. Per ulteriori informazioni sulla modalità di riconfigurazione di un gruppo di istanze e sulle valutazioni da effettuare per tale operazione, consulta [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

Configurazione delle applicazioni per l'uso di una macchina virtuale Java specifica

Le versioni di Amazon EMR hanno versioni predefinite diverse di Java Virtual Machine (JVM). Questa pagina illustra il supporto JVM per diverse versioni e applicazioni.

Considerazioni

- Per le versioni 5.0.0 e successive di Amazon EMR, la Java Virtual Machine (JVM) predefinita è Java 8.

- Per le versioni 6.9.0 e successive di Amazon EMR, Trino supporta Java 17 per impostazione predefinita. Per ulteriori informazioni su Java 17 con Trino, consulta [Aggiornamenti di Trino a Java 17](#) sul blog di Trino.
- Per le versioni 6.12.0 e successive di Amazon EMR, alcune applicazioni supportano anche Java 11 e 17. Per informazioni sulle versioni Java supportate per le applicazioni, consulta la [Guida ai rilasci di Amazon EMR](#). Tieni presente che Amazon EMR supporta solo l'esecuzione di una versione di runtime in un cluster e non supporta l'esecuzione di nodi o applicazioni diversi su versioni di runtime diverse sullo stesso cluster.

Ricorda le seguenti considerazioni specifiche dell'applicazione quando scegli la versione di runtime:

Note sulla configurazione Java per applicazioni specifiche

Applicazione	Note sulla configurazione Java
Spark	<p>Per eseguire Spark con una versione Java non predefinita, devi configurare sia Spark che Hadoop. Per alcuni esempi, consulta Sovrascrittura della JVM.</p> <ul style="list-style-type: none"> • Configura <code>JAVA_HOME</code> in <code>spark-env</code> per aggiornare il runtime Java dei processi dell'istanza primaria. Ad esempio, <code>spark-submit</code>, <code>spark-shell</code> e Spark History Server. • Modifica della configurazione Hadoop per aggiornare il runtime Java degli esecutori Spark e YARN ApplicationMaster
Spark RAPIDS	Puoi eseguire RAPIDS con la versione Java configurata per Spark.
Iceberg	Puoi eseguire Iceberg con la versione Java configurata dell'applicazione che lo utilizza.
Delta	Puoi eseguire Delta con la versione Java configurata dell'applicazione che lo utilizza.

Applicazione	Note sulla configurazione Java
Hudi	Puoi eseguire Hudi con la versione Java configurata dell'applicazione che lo utilizza.
Hive	Per impostare la versione Java su 11 o 17 per Hive, configura l'impostazione JVM di Hadoop sulla versione Java che desideri utilizzare.
HBase	Per aggiornare la JVM per HBase, modifica <code>hbase-env</code> . Per impostazione predefinita, Amazon EMR imposta la JVM di HBase in base alla configurazione JVM per Hadoop, a meno che le impostazioni non vengano sostituite in <code>hbase-env</code> . Per alcuni esempi, consulta Sovrascrittura della JVM .
Flink	Per aggiornare la JVM per Flink, modifica <code>flink-conf</code> . Per impostazione predefinita, Amazon EMR imposta la JVM di Flink in base alla configurazione JVM per Hadoop, a meno che le impostazioni non vengano sostituite in <code>flink-conf</code> . Per ulteriori informazioni, consulta Configurazione di Flink per l'esecuzione con Java 11 .
Oozie	Per configurare Oozie per l'esecuzione su Java 11 o 17, configura Oozie Server, Oozie LauncherAM Launcher AM e modifica le configurazioni degli eseguibili e dei processi sul lato client. Puoi anche configurare <code>EmbeddedOozieServer</code> per l'esecuzione su Java 17. Per ulteriori informazioni, consulta Configurazione della versione Java per Oozie .
Pig	Pig supporta solo Java 8. Non puoi usare Java 11 o 17 con Hadoop ed eseguire Pig sullo stesso cluster.

Sovrascrittura della JVM

Per sostituire l'impostazione JVM per un rilascio di Amazon EMR, ad esempio per utilizzare Java 17 con un cluster che utilizza il rilascio 6.12.0 di Amazon EMR, fornisci l'impostazione JAVA_HOME alla relativa classificazione dell'ambiente, che è *application*-env per tutte le applicazioni eccetto Flink. Per Flink, la classificazione dell'ambiente è *flink-conf*. Per i passaggi per configurare il runtime Java con Flink, consulta [Configurazione di Flink per l'esecuzione con Java 11](#).

Argomenti

- [Sostituisci l'impostazione JVM con Apache Spark](#)
- [Sostituisci l'impostazione JVM con Apache HBase](#)
- [Sostituzione dell'impostazione JVM con Apache Hadoop e Hive](#)

Sostituisci l'impostazione JVM con Apache Spark

Quando usi Spark con Amazon EMR rilascio 6.12 e successivi, se scrivi un driver per l'invio in modalità cluster, il driver utilizza Java 8, ma puoi impostare l'ambiente in modo che gli esecutori utilizzino Java 11 o 17. E quando utilizzi Spark con rilasci di Amazon EMR precedenti a 5.x e scrivi un driver per l'invio in modalità cluster, il driver utilizza Java 7. Tuttavia, puoi impostare l'ambiente per garantire che gli esecutori utilizzino Java 8.

Per sostituire la JVM per Spark, ti consigliamo di impostare entrambe le classificazioni Hadoop e Spark.

```
{
  "Classification": "hadoop-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
          }
        }
      ],
      "Properties": {}
    },
  "Classification": "spark-env",
```

```

    "Configurations": [
      {
"Classification": "export",
        "Configurations": [],
        "Properties": {
"JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
        }
      },
    ],
    "Properties": {}
  }

```

Sostituisci l'impostazione JVM con Apache HBase

Per configurare HBase per l'utilizzo di Java 11, puoi impostare la seguente configurazione all'avvio del cluster.

```

[
  {
    "Classification": "hbase-env",
    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-11"
        }
      }
    ],
    "Properties": {}
  }
]

```

Sostituzione dell'impostazione JVM con Apache Hadoop e Hive

L'esempio seguente mostra come impostare la JVM sulla versione 17 per Hadoop e Hive.

```

[
  {
    "Classification": "hadoop-env",
    "Configurations": [
      {

```



```

        "Classification": "export",
        "Configurations": [],
        "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/jre-17"
        }
    },
    "Properties": {}
}
]

```

Porte di servizio

Di seguito sono elencate le porte di servizio YARN e HDFS. Queste impostazioni riflettono i valori predefiniti di Hadoop. Gli altri servizi applicativi sono ospitati su porte predefinite, a meno che non sia diversamente documentato. Per ulteriori informazioni, consulta la documentazione di progetto dell'applicazione.

Impostazioni porta per YARN e HDFS

Impostazione	Nome host/Porta
fs.default.name	impostazione predefinita (hdfs:// <i>emrDeterminedIP</i> :8020)
dfs.datanode.address	valore di default (0.0.0.0:50010)
dfs.datanode.http.address	valore di default (0.0.0.0:50075)
dfs.datanode.https.address	valore di default (0.0.0.0:50475)
dfs.datanode.ipc.address	valore di default (0.0.0.0:50020)
dfs.http.address	valore di default (0.0.0.0:50070)
dfs.https.address	valore di default (0.0.0.0:50470)
dfs.secondary.http.address	valore di default (0.0.0.0:50090)
yarn.nodemanager.address	valore di default (\${yarn.nodemanager.hostname}:0)

Impostazione	Nome host/Porta
<code>yarn.nodemanager.localizer.address</code>	valore di default (<code>\${yarn.nodemanager.hostname}:8040</code>)
<code>yarn.nodemanager.webapp.address</code>	valore di default (<code>\${yarn.nodemanager.hostname}:8042</code>)
<code>yarn.resourcemanager.address</code>	valore di default (<code>\${yarn.resourcemanager.hostname}:8032</code>)
<code>yarn.resourcemanager.admin.address</code>	valore di default (<code>\${yarn.resourcemanager.hostname}:8033</code>)
<code>yarn.resourcemanager.resource-tracker.address</code>	valore di default (<code>\${yarn.resourcemanager.hostname}:8031</code>)
<code>yarn.resourcemanager.scheduler.address</code>	valore di default (<code>\${yarn.resourcemanager.hostname}:8030</code>)
<code>yarn.resourcemanager.webapp.address</code>	valore di default (<code>\${yarn.resourcemanager.hostname}:8088</code>)
<code>yarn.web-proxy.address</code>	valore di default (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

Note

Il termine *emrDeterminedIP* rappresenta un indirizzo IP generato dal piano di controllo Amazon EMR. Nella versione più recente, questa convenzione è stata rimossa, eccetto che per le impostazioni `yarn.resourcemanager.hostname` e `fs.default.name`.

Utenti delle applicazioni

Le applicazioni eseguono i processi come proprio utente. Ad esempio, le JVM Hive vengono eseguite come utente `hive`, le JVM MapReduce come `mapred` e così via. Questa procedura è illustrata nell'esempio di stato del processo seguente:

```

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_resource manager -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

```

Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR

Puoi utilizzare il repository degli artifact di Amazon EMR per creare il codice processo Apache Hive e Apache Hadoop rispetto alle versioni esatte di librerie e dipendenze disponibili con versioni di rilascio di Amazon EMR specifiche, a partire dalla versione di rilascio 5.18.0 di Amazon EMR. La creazione in base agli artifact di Amazon EMR nel repository consente di evitare problemi di percorso della classe di runtime, garantendo che le versioni delle librerie su cui è creato il processo siano della stessa versione fornita al cluster in fase di runtime. Attualmente, gli artifact di Amazon EMR sono disponibili solo per le build Maven.

Per accedere al repository degli artefatti, aggiungi l'URL del repository al file delle impostazioni Maven o a un file di configurazione `pom.xml` specifico del progetto. Puoi quindi specificare le dipendenze nella configurazione del progetto. Per le versioni delle dipendenze, utilizza la versione elencata in [Versioni dei componenti](#) per il rilascio desiderato [Versioni del rilascio di Amazon EMR 5.x](#). Ad esempio, le versioni dei componenti per il rilascio di Amazon EMR più recente sono disponibili all'indirizzo [the section called "Versioni dei componenti"](#). Se un artefatto del tuo progetto non è elencato in [Versioni dei componenti](#), specifica la versione elencata per Hive e Hadoop in tale rilascio. Ad esempio, per i componenti Hadoop nel rilascio di Amazon EMR 5.18.0, la versione è `2.8.4-amzn-1`.

L'URL del repository degli artefatti presenta la seguente sintassi:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *s3-endpoint* è l'endpoint Amazon Simple Storage Service (Amazon S3) della regione per il repository e *region-ID* è la regione corrispondente. Ad esempio `s3.us-west-1.amazonaws.com` e `us-west-1`. Per ulteriori informazioni, consulta [Endpoint per Amazon S3 in Riferimenti generali di Amazon Web Services](#). Non vi è alcuna differenza negli artefatti tra regioni, perciò è possibile specificare la regione più utile per il proprio ambiente di sviluppo.
- *emr-release-label* è l'etichetta del rilascio del cluster Amazon EMR che eseguirà il codice. Le etichette di rilascio sono nel formato `emr-x.x.x`, ad esempio `emr-5.36.1`. Una serie di release EMR può includere più versioni. Ad esempio, se si utilizza EMR versione 5.24.1, utilizzare la prima etichetta di release EMR all'interno della serie 5.24, `emr-5.24.0`, nell'URL del repository degli artefatti:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

Example Configurazione per Maven pom.xml

L'esempio pom.xml seguente consente di configurare un progetto Maven da creare in base agli artefatti emr-5.18.0 Apache Hadoop e Apache Hive, utilizzando il repository degli artefatti us-west-1. Le versioni degli snapshot non sono disponibili nel repository degli artefatti, perciò gli snapshot sono disabilitati in pom.xml. Le ellissi (...) nell'esempio di seguito indicano omissione di altri parametri di configurazione. Non copiarli nel progetto Maven.

```
<project>
  ...
  <repositories>
    ...
    <repository>
      <id>emr-5.18.0-artifacts</id>
      <name>EMR 5.18.0 Releases Repository</name>
      <releases>
        <enabled>true</enabled>
      </releases>
      <snapshots>
        <enabled>false</enabled>
      </snapshots>
      <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
    </repository>
    ...
  </repositories>
  ...
  <dependencies>
    ...
    <dependency>
      <groupId>org.apache.hive</groupId>
      <artifactId>hive-exec</artifactId>
      <version>2.3.3-amzn-2</version>
    </dependency>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
      <artifactId>hadoop-common</artifactId>
      <version>2.8.4-amzn-1</version>
    </dependency>
  </dependencies>
</project>
```

```
...  
</dependencies>  
  
</project>
```

File system EMR (EMRFS)

Il file system EMR (EMRFS) è un'implementazione di HDFS che tutti i cluster Amazon EMR utilizzano per leggere e scrivere file normali da Amazon EMR direttamente in Amazon S3. EMRFS offre il vantaggio di archiviare dati persistenti in Amazon S3 per l'utilizzo con Hadoop fornendo nel contempo caratteristiche come la crittografia dei dati.

La crittografia dei dati ti consente di crittografare oggetti che EMRFS scrive su Amazon S3 e consente a EMRFS di utilizzare oggetti crittografati in Amazon S3. Se utilizzi Amazon EMR versione 4.8.0 o versioni successive, puoi utilizzare le configurazioni di sicurezza per configurare la crittografia degli oggetti EMRFS in Amazon S3, insieme ad altre impostazioni di crittografia. Per ulteriori informazioni, consulta [Opzioni di crittografia](#). Se utilizzi una versione precedente di Amazon EMR, puoi configurare manualmente le impostazioni di crittografia. Per ulteriori informazioni, consulta [Configurazione della crittografia di Amazon S3 mediante le proprietà di EMRFS](#).

Amazon S3 offre una forte coerenza di lettura dopo scrittura per tutte le operazioni GET, PUT e LIST in tutte le Regioni AWS. Questo significa che ciò che scrivi usando EMRFS è ciò che leggerai da Amazon S3, senza alcun impatto sulle prestazioni. Per maggiori informazioni, consulta il [modello di consistenza dei dati di Amazon S3](#).

Quando utilizzi Amazon EMR versione 5.10.0 o versioni successive, puoi utilizzare ruoli IAM differenti per le richieste EMRFS ad Amazon S3 in funzione degli utenti del cluster, dei gruppi o del percorso dei dati EMRFS in Amazon S3. Per ulteriori informazioni, consulta [Configurazione dei ruoli IAM per le richieste EMRFS a Amazon S3](#).

Warning

Prima di attivare l'esecuzione speculativa per i cluster Amazon EMR che eseguono i processi Apache Spark, consulta le seguenti informazioni.

EMRFS include il committer ottimizzato S3 EMRFS, un'alternativa all'implementazione di OutputCommittee che è ottimizzata per scrivere i file in Amazon S3 quando si usa EMRFS. Se attivi la funzione di esecuzione speculativa di Apache Spark con applicazioni che scrivono dati su Amazon S3 e non utilizzi il committer EMRFS ottimizzato per S3, è possibile che si verifichino problemi di correttezza dei dati descritti in [SPARK-10063](#). Ciò può verificarsi se si utilizzano versioni di Amazon EMR precedenti al rilascio 5.19 o se si scrivono file su Amazon S3 utilizzando formati come ORC e CSV. Questi formati non sono supportati dal committer ottimizzato per EMRFS S3. Per un elenco completo dei requisiti per l'utilizzo del committer EMRFS ottimizzato per S3, consulta [Requisiti per il committer EMRFS ottimizzato per S3](#).

La scrittura diretta EMRFS viene in genere utilizzata quando il committer EMRFS ottimizzato per S3 non è supportato, ad esempio durante la scrittura di quanto segue:

- Un formato di output diverso da Parquet, ad esempio ORC o testo
- File Hadoop che utilizzano l'API Spark RDD.
- Parquet con Hive SerDe. Consulta [Conversione della tabella Parquet con il metastore Hive](#).

La scrittura diretta EMRFS non viene utilizzata nei seguenti scenari:

- Quando il committer ottimizzato S3 EMRFS è abilitato. Consulta [Requisiti per il committer ottimizzato S3 EMRFS](#).
- Quando si scrivono partizioni dinamiche con `partitionOverwriteMode` impostato su `dynamic`.
- Quando si scrive su percorsi di partizione personalizzati, ad esempio percorsi che non sono conformi alla convenzione del percorso della partizione predefinita di Hive.
- Quando si utilizzano file system diversi da EMRFS, come la scrittura su HDFS o l'utilizzo del file system S3A.

Per determinare se la tua applicazione utilizza la scrittura diretta in Amazon EMR 5.14.0 o versioni successive, abilita la registrazione INFO di Spark. Se una riga del log contenente il testo "Direct Write: ENABLED (Scrittura diretta: ABILITATA)" è presente nei log del driver Spark o nei log del container dell'executor Spark, l'applicazione Spark ha scritto utilizzando la scrittura diretta.

Per impostazione predefinita, l'esecuzione speculativa è impostata su OFF su Amazon EMRclusters. Ti consigliamo di non attivare l'esecuzione speculativa se entrambe queste condizioni sono vere:

- Stai scrivendo dati su Amazon S3.
- I dati sono scritti in un formato diverso da Apache Parquet o in formato Apache Parquet che non utilizza il committer EMRFS ottimizzato per S3.

Se attivi l'esecuzione speculativa Spark e scrivi dati su Amazon S3 utilizzando la scrittura diretta EMRFS, potresti riscontrare una perdita di dati intermittente. Quando scrivi dati su HDFS o in Parquet utilizzando il committer EMRFS ottimizzato per S3, Amazon EMR non utilizza la scrittura diretta e questo problema non si verifica.

Se hai bisogno di scrivere dati in formati che utilizzano la scrittura diretta EMRFS da Spark ad Amazon S3 e utilizzi l'esecuzione speculativa, ti consigliamo di scrivere su HDFS e in seguito di trasferire i file di output su Amazon S3 utilizzando S3DistCP.

Argomenti

- [Visualizzazione coerente](#)
- [Autorizzazione di accesso ai dati di EMRFS in Amazon S3](#)
- [Gestione dell'endpoint AWS Security Token Service predefinito](#)
- [Configurazione della crittografia di Amazon S3 mediante le proprietà di EMRFS](#)

Visualizzazione coerente

Warning

Il 1° giugno 2023, la visualizzazione coerente di EMRFS raggiungerà la fine del supporto standard per le future versioni di Amazon EMR. La visualizzazione coerente di EMRFS continuerà a funzionare per le versioni esistenti.

Con il rilascio di una solida coerenza di lettura dopo scrittura di Amazon S3 il 1 dicembre 2020, non è più necessario utilizzare la visualizzazione coerente EMRFS (CV EMRFS) con i cluster Amazon EMR. CV EMRFS è una caratteristica facoltativa che consente ai cluster Amazon EMR di verificare la coerenza a livello di "elenco" e "lettura dopo scrittura" di oggetti Amazon S3. Quando crei un cluster e CV EMRFS è attivato, Amazon EMR crea un database Amazon DynamoDB per archiviare metadati degli oggetti che utilizza per monitorare la coerenza di elenco e lettura dopo la scrittura per gli oggetti S3. Ora puoi disattivare CV EMRFS ed eliminare il database DynamoDB che utilizza, in modo da non accumulare costi aggiuntivi. Le seguenti procedure spiegano come verificare la caratteristica CV, disattivarla ed eliminare il database DynamoDB utilizzato dalla caratteristica.

Verificare se si sta utilizzando la caratteristica CV EMRFS

1. Vai alla scheda Configuration (Configurazione). Se il cluster ha la seguente configurazione, utilizza CV EMRFS.

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. In alternativa, utilizza la AWS CLI per descrivere il cluster con l'[API describe-cluster](#). Se l'output contiene `fs.s3.consistent: true`, il cluster utilizza CV EMRFS.

Disattivare CV EMRFS sui cluster Amazon EMR

Per disattivare la caratteristica CV EMRFS, utilizza una delle seguenti tre opzioni. Ti consigliamo di testare queste opzioni nel tuo ambiente di test prima di applicarle agli ambienti di produzione.

1. Per arrestare il cluster esistente e avviare un nuovo cluster senza le opzioni CV EMRFS.
 - a. Prima di arrestare il cluster, assicurati di eseguire il backup dei dati e di informare gli utenti.
 - b. Per arrestare il cluster, segui le istruzioni riportate nella sezione [Arresto di un cluster](#).
 - c. Se utilizzi la console di Amazon EMR per creare un nuovo cluster, accedi a Advanced Options (Opzioni avanzate). Nella sezione Edit software settings (Modifica impostazioni software), deseleziona l'opzione per attivare CV EMRFS. Se la casella di controllo EMRFS consistent view (Visualizzazione coerente EMRFS) è disponibile, lasciala deselezionata.
 - d. Se utilizzi AWS CLI per creare un nuovo cluster con l'[API create-cluster](#), non utilizzare l'opzione `--emrfs`, che attiva CV EMRFS.
 - e. Se utilizzi un SDK o AWS CloudFormation per creare un nuovo cluster, non utilizzare nessuna delle configurazioni elencate in [Configure consistent view](#) (Configura visualizzazione coerente).
2. Clonare un cluster e rimuovere CV EMRFS
 - a. Nella console di Amazon EMR, scegli il cluster che utilizza CV EMRFS.
 - b. Nella parte superiore della pagina Cluster Details (Dettagli del cluster), fai clic su Clone (Clona).
 - c. Scegli Previous (Precedente) e passa alla Fase 1: software e fasi.
 - d. In Edit software settings (Modifica impostazioni software), rimuovi CV EMRFS. In Edit configuration (Modifica configurazione), elimina le seguenti configurazioni nella classificazione `emrfs-site`. Se stai caricando JSON da un bucket S3, devi modificare l'oggetto S3.

```
[
  {"classification":
    "emrfs-site",
    "properties": {
      "fs.s3.consistent.retryPeriodSeconds": "10",
```

```

    "fs.s3.consistent":"true",
    "fs.s3.consistent.retryCount":"5",
    "fs.s3.consistent.metadata.tableName":"EmrFSMetadata"
  }
}
]

```

3. Rimuovere CV EMRFS da un cluster che utilizza gruppi di istanze

- a. Utilizza il seguente comando per verificare se un singolo cluster EMR utilizza la tabella DynamoDB associata a CV EMRFS o se più cluster condividono la tabella. Il nome della tabella è specificato in `fs.s3.consistent.metadata.tableName`, come descritto in [Configure consistent view](#) (Configura visualizzazione coerente). Il nome predefinito della tabella utilizzato da CV EMRFS è `EmrFSMetadata`.

```

aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName

```

- b. Se il tuo cluster non condivide il database DynamoDB con un altro cluster, utilizza il seguente comando per riconfigurare il cluster e disattivare CV EMRFS. Per ulteriori informazioni, consulta la sezione [Riconfigurazione di un gruppo di istanze in un cluster in esecuzione](#).

```

aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json

```

Questo comando apre il file che desideri modificare. Modifica il file utilizzando le seguenti configurazioni.

```

{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
          "Classification": "emrfs-site",
          "Properties": {
            "fs.s3.consistent": "false"
          }
        },
        "Configurations": []
      ]
    }
  ]
}

```

```
]
}
]
}
```

- c. Se il tuo cluster condivide la tabella DynamoDB con un altro cluster, disattiva CV EMRFS su tutti i cluster quando nessun cluster sta modificando alcun oggetto nella posizione S3 condivisa.

Eliminare le risorse Amazon DynamoDB associate a CV EMRFS

Dopo aver rimosso CV EMRFS dai cluster Amazon EMR, elimina le risorse DynamoDB associate a CV EMRFS. Fino a quando non lo farai, continuerai a sostenere addebiti DynamoDB associati a CV EMRFS.

1. Controlla i parametri di CloudWatch per la tua tabella DynamoDB e verifica che la tabella non sia utilizzata da alcun cluster.
2. Elimina la tabella DynamoDB.

```
aws dynamodb delete-table --table-name <your-table-name>
```

Eliminare le risorse Amazon SQS associate a CV EMRFS

1. Se hai configurato il cluster affinché invii notifiche di incoerenza ad Amazon SQS, puoi eliminare tutte le code SQS.
2. Cerca il nome della coda Amazon SQS specificato in `fs.s3.consistent.notification.SQS.queueName`, come descritto in [Configure consistent view](#) (Configura visualizzazione coerente). Il formato predefinito del nome della coda è `EMRFS-Inconsistency-<j-cluster ID>`.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'
aws sqs delete-queue --queue-url <your-queue-url>
```

Interrompere l'utilizzo della CLI di EMRFS

- La [CLI di EMRFS](#) gestisce i metadati generati da CV EMRFS. Poiché il supporto standard per CV EMRFS terminerà nelle versioni future di Amazon EMR, anche il supporto per la CLI di EMRFS giungerà al termine.

Argomenti

- [Attivazione della visualizzazione coerente](#)
- [Monitoraggio di oggetti in Amazon S3 con la visualizzazione coerente di EMRFS](#)
- [Logica di ripetizione dei tentativi](#)
- [Metadati della visualizzazione coerente EMRFS](#)
- [Configurazione di notifiche di coerenza per CloudWatch e Amazon SQS](#)
- [Configurazione della visualizzazione coerente](#)
- [Riferimento ai comandi CLI di EMRFS](#)

Attivazione della visualizzazione coerente

Puoi abilitare la crittografia lato server di Amazon S3 o la visualizzazione coerente per EMRFS utilizzando la AWS Management Console, la AWS CLI o la classificazione di configurazione `emrfs-site`.

Per configurare la visualizzazione coerente mediante la console

1. Passa alla nuova console Amazon EMR e seleziona **Passa alla vecchia console** dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona **Create cluster (Crea cluster)**, **Go to advanced options (Vai alle opzioni avanzate)**.
3. Scegliere le impostazioni per **Step 1: Software and Steps (Fase 1: software e fasi)** e **Step 2: Hardware (Fase 2: hardware)**.
4. Per **Step 3: General Cluster Settings (Fase 3: impostazioni cluster generali)**, in **Additional Options (Opzioni aggiuntive)**, scegliere **EMRFS consistent view (Visualizzazione coerente EMRFS)**.
5. Per **EMRFS Metadata store (Store di metadati EMRFS)**, digitare il nome dello store di metadati. Il valore predefinito è **EmrFSMetadata**. Se la tabella `EmrFSMetadata` non esiste, viene creata automaticamente in DynamoDB.

Note

Amazon EMR non rimuove automaticamente i metadati EMRFS da DynamoDB quando il cluster viene terminato.

6. In **Number of retries** (Numero di tentativi), digitare un valore intero. Se viene rilevata un'inconsistenza, EMRFS effettua questo numero di tentativi di chiamata di Amazon S3. Il valore predefinito è **5**.
7. Per **Retry period (in seconds)** (Intervallo tra tentativi (in secondi)), digitare un valore intero. Questo è il periodo di tempo durante il quale EMRFS attende tra un tentativo e l'altro. Il valore predefinito è **10**.

Note

I tentativi successivi utilizzano un backoff esponenziale.

Per avviare un cluster con la visualizzazione coerente attivata mediante l'AWS CLI

Si consiglia di installare la versione corrente dell'AWS CLI. Per scaricare la versione più recente, consulta <https://aws.amazon.com/cli/>.

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs  
Consistent=true \  
--release-label emr-5.36.1 --ec2-attributes KeyName=myKey
```

Per verificare se la visualizzazione coerente è attivata utilizzando la AWS Management Console

- Per verificare se la visualizzazione coerente è attivata nella console, accedere a Cluster List (Elenco cluster) e selezionare il nome di cluster per visualizzare Cluster Details (Dettagli cluster). Il campo "EMRFS consistent view" (Visualizzazione coerente EMRFS) ha un valore Enabled o Disabled.

Per verificare se la visualizzazione coerente è attivata esaminando il file **emrfs-site.xml**

- È possibile verificare se la consistenza è attivata, esaminando il file di configurazione `emrfs-site.xml` nel nodo master del cluster. Se il valore booleano per `fs.s3.consistent` è impostato su `true`, la visualizzazione coerente è attivata per le operazioni del file system che implicano Amazon S3.

Monitoraggio di oggetti in Amazon S3 con la visualizzazione coerente di EMRFS

EMRFS crea una visualizzazione coerente degli oggetti in Amazon S3 aggiungendo informazioni su tali oggetti ai metadati EMRFS. EMRFS aggiunge questi elenchi ai relativi metadati quando:

- Un oggetto viene scritto da EMRFS durante un processo Amazon EMR.
- Un oggetto viene sincronizzato con i metadati EMRFS o importato negli stessi mediante la CLI di EMRFS.

Gli oggetti letti da EMRFS non vengono automaticamente aggiunti ai metadati. Quando EMRFS elimina un oggetto, un elenco rimane ancora nei metadati con lo stato eliminato fino a che non viene rimosso mediante la CLI di EMRFS. Per ulteriori informazioni sulla CLI, consulta [Riferimento ai comandi CLI di EMRFS](#). Per ulteriori informazioni sulla rimozione di elenchi nei metadati EMRFS, consulta [Metadati della visualizzazione coerente EMRFS](#).

Per ogni operazione Amazon S3, EMRFS cerca nei metadati informazioni sul set di oggetti nella visualizzazione coerente. Se EMRFS rileva che Amazon S3 è inconsistente durante una di queste operazioni, tenta di nuovo l'operazione in base ai parametri definiti nelle proprietà di configurazione `emrfs-site`. Quando EMRFS ha esaurito i tentativi, genera un'eccezione `ConsistencyException` o registra l'eccezione e continua il flusso di lavoro. Per ulteriori

informazioni sulla logica di ripetizione dei tentativi, consulta [Logica di ripetizione dei tentativi](#). Puoi trovare delle eccezioni `ConsistencyExceptions` nei log, ad esempio:

- `listStatus`: nessun oggetto Amazon S3 per l'elemento di metadati `/S3_bucket/dir/object`
- `getFileStatus`: la chiave `dir/file` è presente nei metadati ma non in Amazon S3

Se elimini direttamente da Amazon S3 un oggetto monitorato con la visualizzazione coerente EMRFS, EMRFS considera quell'oggetto come inconsistente poiché nei metadati è ancora elencato come presente in Amazon S3. Se i metadati non sono più sincronizzati con gli oggetti che EMRFS monitora in Amazon S3, puoi utilizzare il sottocomando `sync` della CLI di EMRFS per ripristinare i metadati affinché corrispondano ad Amazon S3. Per rilevare le discrepanze tra i metadati e Amazon S3, utilizza il comando `diff`. Infine, EMRFS presenta una sola visualizzazione coerente degli oggetti a cui si fa riferimento nei metadati ed è possibile che nello stesso percorso Amazon S3 vi siano altri oggetti che non sono monitorati. Quando EMRFS elenca gli oggetti in un percorso Amazon S3, restituisce il superset degli oggetti monitorati nei metadati e quelli in quel percorso Amazon S3.

Logica di ripetizione dei tentativi

EMRFS tenta di verificare la consistenza degli elenchi per gli oggetti monitorati nei relativi metadati con un determinato numero di tentativi. Il predefinito è 5. In caso di superamento del numero di tentativi consentiti, il processo di origine restituisce un errore a meno che `fs.s3.consistent.throwExceptionOnInconsistency` non sia impostato su `false`, nel qual caso registra solo gli oggetti monitorati come inconsistenti. Per impostazione predefinita, EMRFS utilizza una policy per tentativi di backoff esponenziale, ma è anche possibile impostare una policy fissa. È anche possibile che gli utenti preferiscano riprovare per un determinato periodo di tempo prima di procedere con il resto del processo senza generare un'eccezione. A questo proposito, è necessario impostare `fs.s3.consistent.throwExceptionOnInconsistency` su `false`, `fs.s3.consistent.retryPolicyType` su `fixed` e `fs.s3.consistent.retryPeriodSeconds` sul valore desiderato. L'esempio seguente crea un cluster con la consistenza attivata, che registra le inconsistenze e imposta un intervallo fisso di 10 secondi tra tentativi:

Example Impostazione di un intervallo fisso tra tentativi

```
aws emr create-cluster --release-label emr-5.36.1 \  
--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

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Per ulteriori informazioni, consulta [Visualizzazione coerente](#).

Le configurazioni EMRFS per IMDS ricevono chiamate regionali

EMRFS si affida all'IMDS (servizio di metadati dell'istanza) per ottenere la Regione dell'istanza e gli endpoint di Amazon S3, DynamoDB o AWS KMS. Tuttavia, IMDS ha un limite al numero di richieste che può gestire e le richieste che superano tale limite falliranno. Questo limite IMDS può causare l'inizializzazione di fallimenti EMRFS e causare il fallimento della query o del comando. È possibile utilizzare il seguente meccanismo di riprova di backoff esponenziale randomizzato e le proprietà di configurazione di una Regione di fallback in `emrfs-site.xml` per risolvere lo scenario in cui tutti i tentativi falliscono.

```
<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>
```

Metadati della visualizzazione coerente EMRFS

La visualizzazione coerente EMRFS verifica la consistenza utilizzando una tabella DynamoDB per monitorare gli oggetti in Amazon S3 che sono stati sincronizzati o creati da EMRFS. I metadati sono utilizzati per monitorare tutte le operazioni (lettura, scrittura, aggiornamento e copia) e alcun contenuto effettivo è memorizzato negli stessi. Questi metadati sono utilizzati per confermare che gli oggetti o i metadati ricevuti da Amazon S3 corrispondono a quanto previsto. Tale conferma consente a EMRFS di verificare la consistenza a livello di elenco e read-after-write dei nuovi oggetti che scrive su Amazon S3 o degli oggetti che sincronizza. Più cluster possono condividere gli stessi metadati.

Come aggiungere voci ai metadati

Puoi utilizzare i sottocomandi `sync` o `import` per aggiungere voci ai metadati. `sync` riflette lo stato degli oggetti Amazon S3 in un percorso, mentre `import` viene utilizzato esclusivamente per aggiungere nuove voci ai metadati. Per ulteriori informazioni, consulta [Riferimento ai comandi CLI di EMRFS](#).

Come verificare le differenze tra metadati e oggetti in Amazon S3

Per verificare le differenze tra i metadati e Amazon S3, utilizza il sottocomando `diff` della CLI di EMRFS. Per ulteriori informazioni, consulta [Riferimento ai comandi CLI di EMRFS](#).

Come determinare se le operazioni di metadati sono sottoposte a throttling

EMRFS imposta i limiti di capacità di velocità effettiva di default sui metadati per le operazioni di lettura e scrittura rispettivamente su 500 e 100 unità. Con un gran numero di oggetti o bucket, le operazioni possono superare questa capacità e quindi essere sottoposte a limitazione da DynamoDB. Ad esempio, un'applicazione può portare EMRFS a generare un'eccezione `ProvisionedThroughputExceededException` se si sta eseguendo un'operazione che supera questi limiti di capacità. Con la limitazione, la CLI di EMRFS esegue un nuovo tentativo di scrittura sulla tabella DynamoDB utilizzando il [backoff esponenziale](#) fino al termine dell'operazione o fino a quando non viene raggiunto il numero massimo di tentativi di scrittura di oggetti da Amazon EMR su Amazon S3.

È possibile configurare i propri limiti di capacità di velocità effettiva. Tuttavia, DynamoDB dispone di limiti di partizione rigorosi di 3000 unità di capacità in lettura (RCU) e 1000 unità di capacità in

scrittura (WCU) al secondo per operazioni di lettura e scrittura. Per evitare errori sync causati dalla limitazione, consigliamo di limitare la velocità effettiva per le operazioni di lettura a meno di 3000 RCU e le operazioni di scrittura a meno di 1000 WCU. Per istruzioni sull'impostazione dei limiti di capacità di velocità effettiva personalizzati, consulta [Configurazione della visualizzazione coerente](#).

Puoi anche visualizzare i parametri di Amazon CloudWatch per i metadati EMRFS nella console di DynamoDB in cui è indicato il numero di richieste di lettura e scrittura sottoposte a limitazione. Se il numero di richieste sottoposte a throttling è diverso da zero, la tua applicazione può potenzialmente beneficiare di una capacità di throughput allocata superiore per le operazioni di lettura o scrittura. Puoi anche ottenere un miglioramento delle prestazioni se le tue operazioni si avvicinano alla capacità di throughput allocata massima per le letture o le scritture durante un periodo di tempo prolungato.

Caratteristiche di throughput per operazioni EMRFS significative

L'impostazione predefinita per le operazioni di lettura e scrittura è rispettivamente di 400 e 100 unità di capacità di throughput. Le caratteristiche prestazionali seguenti offrono un'idea di quale throughput è richiesto per determinate operazioni. Questi test sono stati eseguiti utilizzando un cluster `m3.large` a nodo singolo. Tutte le operazioni erano a thread singolo. Le prestazioni differiscono notevolmente a seconda delle specifiche caratteristiche dell'applicazione e potrebbe essere necessario sperimentare per ottimizzare le operazioni del file system.

Operazione	Media letture al secondo	Media scritture al secondo
create (oggetto)	26.79	6.70
delete (oggetto)	10.79	10.79
delete (directory contenente 1.000 oggetti)	21.79	338.40
getFileStatus (oggetto)	34.70	0
getFileStatus (directory)	19.96	0
listStatus (directory contenente 1 oggetto)	43.31	0

Operazione	Media letture al secondo	Media scritture al secondo
listStatus (directory contenente 10 oggetti)	44.34	0
listStatus (directory contenente 100 oggetti)	84.44	0
listStatus (directory contenente 1.000 oggetti)	308.81	0
listStatus (directory contenente 10.000 oggetti)	416.05	0
listStatus (directory contenente 100.000 oggetti)	823.56	0
listStatus (directory contenente 1 milione di oggetti)	882.36	0
mkdir (continua per 120 secondi)	24.18	4.03
mkdir	12.59	0
rename (oggetto)	19.53	4.88
rename (directory contenente 1.000 oggetti)	23.22	339.34

Per inviare una fase che rimuove dati obsoleti dallo store di metadati

È possibile che gli utenti vogliano rimuovere specifiche voci nei metadati basati su DynamoDB. Ciò può contribuire a ridurre i costi di storage associati alla tabella. Gli utenti hanno la possibilità di rimuovere manualmente o a livello di codice determinate voci utilizzando il sottocomando `delete` della CLI di EMRFS. Tuttavia, se elimini delle voci dai metadati, EMRFS non esegue più alcuna verifica di consistenza.

La rimozione a livello di codice dopo il completamento di un processo può essere eseguita inviando una fase finale al cluster, il quale esegue un comando sulla CLI di EMRFS. Ad esempio, digitare il comando seguente per inviare una fase al cluster allo scopo di eliminare tutte le voci di più di due giorni.

```
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"
  ]
}
```

Utilizzare il valore StepId restituito per esaminare il risultato dell'operazione nei log.

Configurazione di notifiche di coerenza per CloudWatch e Amazon SQS

Puoi attivare i parametri di CloudWatch e i messaggi di Amazon SQS in EMRFS per problemi di consistenza finale in Amazon S3.

CloudWatch

Quando i parametri di CloudWatch sono attivati, un parametro denominato Inconsistency (Incoerenza) viene trasmesso ogni volta che una chiamata API FileSystem non riesce a causa della consistenza finale di Amazon S3.

Visualizzazione dei parametri di CloudWatch per i problemi di consistenza finale di Amazon S3

Per visualizzare il parametro Inconsistency (Incoerenza) nella console di CloudWatch, seleziona i parametri di EMRFS e, in seguito, seleziona una coppia JobFlowId/Metric Name (Nome parametro). Ad esempio: *j-162XXXXXXM2CU ListStatus*, *j-162XXXXXXM2CU GetFileStatus* e così via.

1. Aprire la console CloudWatch all'indirizzo <https://console.aws.amazon.com/cloudwatch/>.
2. In Dashboard (Pannello di controllo), nella sezione Metrics (Parametri), scegliere EMRFS.
3. Nel riquadro Job Flow Metrics (Parametri flusso di elaborazione), selezionare una o più coppie JobFlowId/Metric Name (Nome parametro). Una rappresentazione grafica dei parametri è visualizzata nella finestra illustrata di seguito.

Amazon SQS

Quando le notifiche di Amazon SQS sono attivate, una coda Amazon SQS denominata `EMRFS-Inconsistency-<jobFlowId>` viene creata all'inizializzazione di EMRFS. I messaggi di Amazon SQS sono trasmessi alla coda quando una chiamata API `FileSystem` non riesce a causa della consistenza finale di Amazon S3. Il messaggio contiene informazioni quali `JobFlowId`, `API`, un elenco di percorsi inconsistenti, una traccia di stack e così via. I messaggi possono essere letti utilizzando la console di Amazon SQS o il comando `read-sqs` di EMRFS.

Gestione dei messaggi di Amazon SQS per i problemi di consistenza finale di Amazon S3

È possibile leggere i messaggi Amazon SQS per i problemi di consistenza finale di Amazon S3 mediante la CLI di EMRFS. Per leggere i messaggi da una coda Amazon SQS di EMRFS, digita il comando `read-sqs` e specifica un percorso di output nel file system locale del nodo master per il file di output risultante.

È anche possibile eliminare una coda Amazon SQS di EMRFS utilizzando il comando `delete-sqs`.

1. Per leggere i messaggi da una coda Amazon SQS, digita il comando seguente. Sostituisci *queuename* con il nome della coda Amazon SQS configurata e sostituisci */path/filename* con il percorso del file di output:

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

Ad esempio, per leggere i messaggi di output di Amazon SQS dalla coda predefinita, digita:

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

Note

È inoltre possibile utilizzare le scelte rapide `-q` e `-o` anziché `--queue-name` e `--output-file`.

2. Per eliminare una coda Amazon SQS, digita il comando seguente:

```
emrfs delete-sqs --queue-name queuename
```

Ad esempio, per eliminare la coda predefinita, digitare:

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXXXM2CU
```

Note

È anche possibile utilizzare la scelta rapida `-q` anziché `--queue-name`.

Configurazione della visualizzazione coerente

Puoi configurare ulteriori impostazioni per la visualizzazione coerente fornendole mediante le proprietà di configurazione `emrfs-site`. Ad esempio, puoi scegliere una velocità effettiva DynamoDB predefinita differente fornendo gli argomenti seguenti all'opzione della CLI `--emrfs`, utilizzando la classificazione di configurazione `emrfs-site` (solo Amazon EMR versione 4.x e versioni successive) o un'operazione di bootstrap per configurare il file `emrfs-site.xml` sul nodo master:

Example Modifica dei valori di lettura e scrittura dei metadati di default all'avvio del cluster

```
aws emr create-cluster --release-label emr-5.36.1 --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
```

In alternativa, puoi utilizzare il file di configurazione seguente e salvarlo in locale o in Amazon S3:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.consistent.metadata.read.capacity": "600",
      "fs.s3.consistent.metadata.write.capacity": "300"
    }
  }
]
```

Utilizza la configurazione creata con la sintassi seguente:

```
aws emr create-cluster --release-label emr-5.36.1 --applications Name=Hive \
--instance-type m5.xlarge --instance-count 2 --configurations file:///./myConfig.json
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Le opzioni seguenti possono essere impostate utilizzando delle configurazioni o gli argomenti `--emrfs` di AWS CLI. Per informazioni su questi argomenti, consulta la [Guida di riferimento ai comandi della AWS CLI](#).

Proprietà **`emrfs-site.xml`** per la visualizzazione coerente

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.consistent</code>	false	Quando impostata su true , questa proprietà configura EMRFS per utilizzare DynamoDB per fornire consistenza.
<code>fs.s3.consistent.retryPolicyType</code>	exponential	Questa proprietà identifica la policy da utilizzare quando si effettua un nuovo tentativo per i problemi di consistenza. Le opzioni includono: <code>exponential</code> , <code>fixed</code> o <code>none</code> .
<code>fs.s3.consistent.retryPeriodSeconds</code>	1	Questa proprietà imposta l'intervallo di attesa tra i tentativi di consistenza.
<code>fs.s3.consistent.retryCount</code>	10	Questa proprietà imposta il numero massimo di tentativi quando viene rilevata un'inconsistenza.

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.consistent.throwExceptionOnInconsistency</code>	true	Questa proprietà determina se generare o registrare un'eccezione di consistenza. Quando impostata su true , viene generata un'eccezione <code>ConsistencyException</code> .
<code>fs.s3.consistent.metadata.autoCreate</code>	true	Quando impostata su true , questa proprietà attiva la creazione automatica delle tabelle di metadati.
<code>fs.s3.consistent.metadata.etagVerificationEnabled</code>	true	Con Amazon EMR 5.29.0, questa proprietà è abilitata per impostazione predefinita. Quando abilitato, EMRFS utilizza gli ETag S3 per verificare che gli oggetti letti siano la versione più recente disponibile. Questa caratteristica è utile per i casi d'uso read-after-update in cui i file in S3 vengono sovrascritti mantenendo lo stesso nome. Questa caratteristica di verifica ETag attualmente non funziona con S3 Select.
<code>fs.s3.consistent.metadata.tableName</code>	EmrFSMetadata	Questa proprietà specifica il nome della tabella di metadati in DynamoDB.

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.consistent.metadata.read.capacity</code>	500	Questa proprietà specifica la capacità di lettura di DynamoDB per effettuare il provisioning quando viene creata la tabella di metadati.
<code>fs.s3.consistent.metadata.write.capacity</code>	100	Questa proprietà specifica la capacità di scrittura di DynamoDB per effettuare il provisioning quando viene creata la tabella di metadati.
<code>fs.s3.consistent.fastList</code>	true	Quando impostata su true , questa proprietà utilizza più thread per elencare una directory (quando necessario). La consistenza deve essere attivata per utilizzare questa proprietà.
<code>fs.s3.consistent.fastList.prefetchMetadata</code>	false	Quando impostata su true , questa proprietà attiva il recupero dei metadati per le directory contenenti più di 20.000 elementi.
<code>fs.s3.consistent.notification.CloudWatch</code>	false	Quando impostati su true , i parametri di CloudWatch vengono attivati per le chiamate API FileSystem che non vanno a buon fine a causa di problemi di consistenza finale di Amazon S3.

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.consistent.notification.SQS</code>	false	Quando impostate su true , le notifiche di consistenza finale sono trasmesse a una coda Amazon SQS.
<code>fs.s3.consistent.notification.SQS.queueName</code>	EMRFS-Inconsistency- <jobFlowId>	La modifica di questa proprietà consente di specificare il nome della coda SQS per i messaggi relativi ai problemi di consistenza finale di Amazon S3.
<code>fs.s3.consistent.notification.SQS.customMsg</code>	none	Questa proprietà consente di specificare informazioni personalizzate incluse in messaggi SQS relativi ai problemi di consistenza finale di Amazon S3. Se un valore non è specificato per questa proprietà, il campo corrispondente nel messaggio è vuoto.
<code>fs.s3.consistent.dynamodb.endpoint</code>	none	Questa proprietà consente di specificare un endpoint DynamoDB personalizzato per i metadati della visualizzazione coerente.
<code>fs.s3.useRequesterPaysHeader</code>	false	Quando è impostato su true , questa proprietà consente alle richieste Amazon S3 ai bucket con l'opzione dell'entità pagante delle richieste abilitata.

Riferimento ai comandi CLI di EMRFS

Per impostazione predefinita, la CLI di EMRFS è installata su tutti i nodi master del cluster creati utilizzando Amazon EMR versione 3.2.1 o versioni successive. Puoi utilizzare la CLI di EMRFS per gestire i metadati per la visualizzazione coerente.

Note

Il comando `emrfs` è supportato solo con l'emulazione di terminale VT100. Tuttavia, può funzionare anche con altre modalità emulatore di terminale.

Commando di livello superiore `emrfs`

Il comando di livello superiore `emrfs` supporta la struttura seguente.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
 list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Specifica [opzioni], con o senza [argomenti] come descritto nella tabella seguente. Per [opzioni] specifiche a sottocomandi (`describe-metadata`, `set-metadata-capacity`, ecc.), consulta ogni sottocomando elencato di seguito.

[Opzioni] per `emrfs`

Opzione	Descrizione	Obbligato
<code>-a <i>AWS_ACCESS_KEY_ID</i></code> <code>--access-key <i>AWS_ACCESS_KEY_ID</i></code>	Le chiavi di accesso AWS che utilizzi per scrivere oggetti su Amazon S3 e per creare o accedere a store di metadati in DynamoDB. Per impostazione predefinita, <i>AWS_ACCESS_KEY_ID</i> è impostata sulla chiave di accesso utilizzata per creare il cluster.	No
	La chiave segreta AWS associata alla chiave di accesso che utilizzi per scrivere oggetti su Amazon	No

Opzione	Descrizione	Obbligatorio
<code>-s <i>AWS_SECRET_ACCESS_KEY</i> --secret-key <i>AWS_SECRET_ACCESS_KEY</i></code>	S3 e per creare o accedere a store di metadati in DynamoDB. Per impostazione predefinita, <i>AWS_SECRET_ACCESS_KEY</i> è impostata sulla chiave segreta associata alla chiave di accesso utilizzata per creare il cluster.	
<code>-v --verbose</code>	Fornisce un output dettagliato.	No
<code>-h --help</code>	Visualizza il messaggio di aiuto per il comando <code>emrfs</code> con un'istruzione di utilizzo.	No

Sottocomando `emrfs describe-metadata`

[Opzioni] per `emrfs describe-metadata`

Opzione	Descrizione	Obbligatorio
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	No

Example Esempio di `emrfs describe-metadata`

L'esempio seguente descrive la tabella di metadati di default.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 12
```

Sottocomando `emrfs set-metadata-capacity`

[Opzioni] per `emrfs set-metadata-capacity`

Opzione	Descrizione	Obbligato
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	No
<code>-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i></code>	La capacità di throughput di lettura necessaria per la tabella di metadati. Se l'argomento <i>READ_CAPACITY</i> non è fornito, il valore predefinito è 400.	No
<code>-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i></code>	La capacità di throughput di scrittura necessaria per la tabella di metadati. Se l'argomento <i>WRITE_CAPACITY</i> non è fornito, il valore predefinito è 100.	No

Example Esempio di `emrfs set-metadata-capacity`

L'esempio seguente imposta la capacità di throughput di lettura su 600 e la capacità di scrittura su 150 per una tabella di metadati denominata `EmrMetadataAlt`.

```
$ emrfs set-metadata-capacity --metadata-name EmrMetadataAlt --read-capacity 600 --
write-capacity 150
  read-capacity: 400
  write-capacity: 100
  status: UPDATING
  approximate-item-count (6 hour delay): 0
```

Sottocomando emrfs delete-metadata

[Opzioni] per emrfs delete-metadata

Opzione	Descrizione	Obbligatorio
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è EmrFSMetadata .	No

Example Esempio di emrfs delete-metadata

L'esempio seguente elimina la tabella di metadati di default.

```
$ emrfs delete-metadata
```

Sottocomando emrfs create-metadata

[Opzioni] per emrfs create-metadata

Opzione	Descrizione	Obbligatorio
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è EmrFSMetadata .	No
-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i>	La capacità di throughput di lettura necessaria per la tabella di metadati. Se l'argomento <i>READ_CAPACITY</i> non è fornito, il valore predefinito è 400.	No
		No

Opzione	Descrizione	Obbligatorio
-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i>	La capacità di throughput di scrittura necessaria per la tabella di metadati. Se l'argomento <i>WRITE_CAPACITY</i> non è fornito, il valore predefinito è 100.	

Example Esempio di emrfs create-metadata

L'esempio seguente crea una tabella di metadati denominata EmrFSMetadataAlt.

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

Sottocomando emrfs list-metadata-stores

Il sottocomando emrfs list-metadata-stores non ha [opzioni].

Example Esempio di list-metadata-store

L'esempio seguente elenca le tabelle di metadati.

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

Sottocomando emrfs diff

[Opzioni] per emrfs diff

Opzione	Descrizione	Obbligatorio
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_</i>	No

Opzione	Descrizione	Obbligato
	<i>NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	
<i>s3://s3Path</i>	Il percorso al bucket Amazon S3 da comparare alla tabella di metadati. Sincronizzazione ricorsiva dei bucket.	Sì

Example Esempio di emrfs diff

L'esempio seguente compara la tabella di metadati di default a un bucket Amazon S3.

```
$ emrfs diff s3://elasticmapreduce/samples/cloudfront
BOTH | MANIFEST ONLY | S3 ONLY
DIR elasticmapreduce/samples/cloudfront
DIR elasticmapreduce/samples/cloudfront/code/
DIR elasticmapreduce/samples/cloudfront/input/
DIR elasticmapreduce/samples/cloudfront/logprocessor.jar
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-14.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-15.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-16.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-17.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-18.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-19.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-20.WxYz1234
DIR elasticmapreduce/samples/cloudfront/code/cloudfront-loganalyzer.tgz
```

Sottocomando emrfs delete

[Opzioni] per emrfs delete

Opzione	Descrizione	Obbligato
	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_</i>	No

Opzione	Descrizione	Obbligatorio
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	
<code><i>s3://s3Path</i></code>	Il percorso al bucket Amazon S3 monitorato per la visualizzazione coerente. Sincronizzazione ricorsiva dei bucket.	Sì
<code>-t <i>TIME</i> --time <i>TIME</i></code>	La data di scadenza (interpretata utilizzando l'argomento dell'unità di tempo). Tutte le voci di metadati più vecchie dell'argomento <i>TIME</i> sono eliminate per il bucket specificato.	
<code>-u <i>UNIT</i> --time-unit <i>UNIT</i></code>	La misura utilizzata per interpretare l'argomento di tempo (nanosecondi, microsecondi, millisecondi, secondi, minuti, ore o giorni). Se nessun argomento è specificato, il valore di default è <code>days</code> .	
<code>--read-consumption <i>READ_CONSUMPTION</i></code>	La quantità richiesta di throughput di lettura disponibile utilizzata per l'operazione delete. Se l'argomento <i>READ_CONSUMPTION</i> non è specificato, il valore predefinito è <code>400</code> .	No
<code>--write-consumption <i>WRITE_CONSUMPTION</i></code>	La quantità richiesta di throughput di scrittura disponibile utilizzata per l'operazione delete. Se l'argomento <i>WRITE_CONSUMPTION</i> non è specificato, il valore predefinito è <code>100</code> .	No

Example Esempio di emrfs delete

L'esempio seguente rimuove tutti gli oggetti in un bucket Amazon S3 dai metadati di monitoraggio per la visualizzazione coerente.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

Sottocomando emrfs import

[Opzioni] per emrfs import

Opzione	Descrizione	Obbligatorio
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	No
<i>s3://s3Path</i>	Il percorso al bucket Amazon S3 monitorato per la visualizzazione coerente. Sincronizzazione ricorsiva dei bucket.	Si
--read-consumption <i>READ_CONSUMPTION</i>	La quantità richiesta di throughput di lettura disponibile utilizzata per l'operazione delete. Se l'argomento <i>READ_CONSUMPTION</i> non è specificato, il valore predefinito è 400.	No
--write-consumption <i>WRITE_CONSUMPTION</i>	La quantità richiesta di throughput di scrittura disponibile utilizzata per l'operazione delete. Se l'argomento <i>WRITE_CONSUMPTION</i> non è specificato, il valore predefinito è 100.	No

Example Esempio di emrfs import

L'esempio seguente importa tutti gli oggetti in un bucket Amazon S3 con i metadati di monitoraggio per la visualizzazione coerente. Tutte le chiavi sconosciute vengono ignorate.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

Sottocomando emrfs sync

[Opzioni] per emrfs sync

Opzione	Descrizione	Obbligatorio
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> è il nome della tabella di metadati DynamoDB. Se l'argomento <i>METADATA_NAME</i> non è fornito, il valore predefinito è <code>EmrFSMetadata</code> .	No
<i>s3://s3Path</i>	Il percorso al bucket Amazon S3 monitorato per la visualizzazione coerente. Sincronizzazione ricorsiva dei bucket.	Si
--read-consumption <i>READ_CONSUMPTION</i>	La quantità richiesta di throughput di lettura disponibile utilizzata per l'operazione delete. Se l'argomento <i>READ_CONSUMPTION</i> non è specificato, il valore predefinito è 400.	No
--write-consumption <i>WRITE_CONSUMPTION</i>	La quantità richiesta di throughput di scrittura disponibile utilizzata per l'operazione delete. Se l'argomento <i>WRITE_CONSUMPTION</i> non è specificato, il valore predefinito è 100.	No

Example Esempio di comando emrfs sync

L'esempio seguente importa tutti gli oggetti in un bucket Amazon S3 con i metadati di monitoraggio per la visualizzazione coerente. Tutte le chiavi sconosciute vengono eliminate.

```
$ 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
```

Sottocomando emrfs read-sqs

[Opzioni] per emrfs read-sqs

Opzione	Descrizione	Obbligato
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> è il nome della coda Amazon SQS configurata in <code>emrfs-site.xml</code> . Il valore predefinito è EMRFS-Inconsistency-<code><jobFlowId></code> .	Sì
<code>-o <i>OUTPUT_FILE</i> --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> è il percorso al file di output nel file system locale del nodo master. I messaggi letti dalla coda sono scritti su questo file.	Sì

Sottocomando emrfs delete-sqs

[Opzioni] per emrfs delete-sqs

Opzione	Descrizione	Obbligatorio
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> è il nome della coda Amazon SQS configurata in <code>emrfs-site.xml</code> . Il valore predefinito è EMRFS-Inconsistency-<code><jobFlowId></code> .	Sì

Invio di comandi della CLI di EMRFS come fasi

L'esempio seguente mostra come utilizzare l'utilità `emrfs` sul nodo master mediante AWS CLI o l'API `command-runner.jar` per eseguire il comando `emrfs` come fase. L'esempio utilizza il AWS SDK for Python (Boto3) per aggiungere una fase a un cluster che aggiunge oggetti in un bucket Amazon S3 alla tabella di metadati EMRFS di default.

```
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()
```

Puoi utilizzare il valore `step_id` restituito per verificare il risultato dell'operazione nei log.

Autorizzazione di accesso ai dati di EMRFS in Amazon S3

Per impostazione predefinita, il ruolo EMR per EC2 determina le autorizzazioni per l'accesso ai dati di EMRFS in Amazon S3. Le policy IAM collegate a questo ruolo vengono applicate indipendentemente dall'utente o dal gruppo che effettua la richiesta tramite EMRFS. Il valore predefinito è `EMR_EC2_DefaultRole`. Per ulteriori informazioni, consulta [Ruolo di servizio per le istanze EC2 del cluster \(profilo dell'istanza EC2\)](#).

A partire da Amazon EMR versione 5.10.0, puoi utilizzare una configurazione di sicurezza per specificare ruoli IAM per EMRFS. Ciò ti consente di personalizzare le autorizzazioni per le richieste EMRFS a Amazon S3 per cluster che dispongono di più utenti. Puoi specificare ruoli IAM differenti

per differenti utenti e gruppi e per percorsi di bucket Amazon S3 differenti in base al prefisso in Amazon S3. Quando EMRFS effettua una richiesta ad Amazon S3 che corrisponde a utenti, gruppi o percorsi che specifichi, il cluster utilizza il ruolo corrispondente da te indicato anziché il ruolo EMR per EC2. Per ulteriori informazioni, consulta [Configurazione dei ruoli IAM per le richieste EMRFS a Amazon S3](#).

In alternativa, se le richieste della tua soluzione Amazon EMR non sono soddisfatte dai ruoli IAM per EMRFS, puoi definire una classe di provider di credenziali personalizzato che consente di personalizzare l'accesso ai dati di EMRFS in Amazon S3.

Creazione di un provider di credenziali personalizzato per i dati di EMRFS in Amazon S3

Per creare un provider di credenziali personalizzato, devi implementare le classi [AWSCredentialsProvider](#) e [Configurable](#) di Hadoop.

Per una descrizione dettagliata di questo approccio, consulta [Securely Analyze Data from Another AWS Account with EMRFS \(Analizza i dati in modo sicuro da un altro account AWS con EMRFS\)](#) nel blog sui Big Data AWS. Il post del blog include un tutorial che descrive il processo end-to-end, dalla creazione di ruoli IAM all'avvio del cluster. Fornisce inoltre un esempio di codice Java che implementa la classe di provider di credenziali personalizzato.

I passaggi di base sono i seguenti:

Per specificare un provider di credenziali personalizzato

1. Creare una classe di provider di credenziali personalizzato compilata come file JAR.
2. Eseguire uno script come un'operazione di bootstrap per copiare il file JAR del provider di credenziali personalizzato nella posizione `/usr/share/aws/emr/emrfs/auxlib` sul nodo master del cluster. Per ulteriori informazioni sulle operazioni di bootstrap, consulta l'argomento relativo alla [Creazione di operazioni di bootstrap per installare software aggiuntivi](#).
3. Personalizzare la classificazione `emrfs-site` per specificare la classe che si implementa nel file JAR. Per ulteriori informazioni su come specificare oggetti di configurazione per personalizzare le applicazioni, consulta l'argomento relativo alla [Configurazione di applicazioni](#) nella Guida ai rilasci di Amazon EMR.

L'esempio seguente illustra un comando `create-cluster` che avvia un cluster Hive con parametri di configurazione comuni e include inoltre:

- Un'operazione di bootstrap che esegue lo script, *copy_jar_file.sh*, salvato in *mybucket* in Amazon S3.
- Una classificazione *emrfs-site* che specifica un provider di credenziali personalizzato definito nel file JAR come *MyCustomCredentialsProvider*.

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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-5.36.1 \
--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":[]}]'
```

Gestione dell'endpoint AWS Security Token Service predefinito

EMRFS utilizza il AWS Security Token Service (STS) per recuperare le credenziali di sicurezza temporanee per accedere alle risorse AWS. Le versioni precedenti di Amazon EMR inviano tutte le richieste AWS STS a un singolo endpoint globale all'indirizzo <https://sts.amazonaws.com>. Le versioni 5.31.0, 6.1.0 e versioni successive di Amazon EMR inviano richieste a endpoint AWS STS regionali. Questo riduce la latenza e migliora la validità del token di sessione. Per ulteriori informazioni sugli endpoint AWS STS, consulta [Gestione di AWS STS in una Regione AWS](#) nella Guida per l'utente di AWS Identity and Access Management.

Quando utilizzi la versione 5.31.0 e 6.1.0 e versioni successive di Amazon EMR, puoi sovrascrivere l'endpoint AWS STS predefinito. A questo scopo, devi modificare la proprietà `fs.s3.sts.endpoint` nella tua configurazione `emrfs-site`.

Il seguente esempio AWS CLI imposta l'endpoint AWS STS predefinito utilizzato da EMRFS per l'endpoint globale.

```
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

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

In alternativa, è possibile creare un file di configurazione JSON utilizzando l'esempio seguente e specificarlo utilizzando l'argomento `--configurations` di `emr create-cluster`. Per ulteriori informazioni sull'utilizzo di `--configurations`, consulta la [Guida di riferimento ai comandi della AWS CLI](#).

```
[  
  {  
    "classification": "emrfs-site",  
    "properties": {  
      "fs.s3.sts.endpoint": "https://sts.amazonaws.com"  
    }  
  }  
]
```

Configurazione della crittografia di Amazon S3 mediante le proprietà di EMRFS

Important

A partire da Amazon EMR versione 4.8.0, puoi utilizzare le configurazioni di sicurezza per applicare impostazioni di crittografia più facilmente e con più opzioni. Ti consigliamo di

utilizzare configurazioni di sicurezza. Per ulteriori informazioni, consulta [Configurazione della crittografia dei dati](#). Le istruzioni relative alla console descritte in questa sezione sono disponibili per le versioni precedenti alla versione 4.8.0. Se utilizzi la AWS CLI per configurare la crittografia di Amazon S3 nella configurazione del cluster e in una configurazione di sicurezza nelle versioni successive, la configurazione di sicurezza sostituisce la configurazione del cluster.

Quando crei un cluster, puoi specificare la crittografia lato server (SSE) oppure la crittografia lato client (CSE) per i dati di EMRFS in Amazon S3 utilizzando la console o le proprietà di classificazione `emrfs-site` mediante la AWS CLI o l'SDK EMR. Le crittografie SSE e CSE di Amazon S3 si escludono a vicenda; puoi scegliere una delle due, ma non entrambe.

Per le istruzioni relative all'AWS CLI, consulta la sezione appropriata per il tipo di crittografia utilizzato esposta di seguito.

Per impostare opzioni di crittografia di EMRFS mediante la AWS Management Console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Release (Versione) scegliere la versione 4.7.2 o una versione precedente.
4. In Software and Steps (Software e fasi), scegliere altre opzioni appropriate per l'applicazione, quindi scegliere Next (Avanti).
5. Scegliere le impostazioni nei riquadri Hardware e General Cluster Settings (Impostazioni cluster generali) come appropriato per l'applicazione.
6. Nel riquadro Security (Sicurezza), in Authentication and encryption (Autenticazione e crittografia), selezionare l'opzione S3 Encryption (with EMRFS) (Crittografia S3 (con EMRFS)) da utilizzare.

Note

L'opzione S3 server-side encryption with KMS Key Management (Crittografia lato server S3 con KMS Key Management) (SSE-KMS) non è disponibile quando si utilizza Amazon EMR versione 4.4 o versioni precedenti.

- Se si sceglie un'opzione che utilizza AWS Key Management, scegliere un ID di chiave in AWS KMS Key ID (ID chiave AWS KMS). Per ulteriori informazioni, consulta [Utilizzo di AWS KMS keys per la crittografia EMRFS](#).
 - Se si sceglie S3 client-side encryption with custom materials provider (Crittografia lato client S3 con provider di materiali personalizzato), specificare il nome della classe in Class name (Nome classe) e la posizione del JAR in JAR location (Percorso JAR). Per ulteriori informazioni, consulta [File crittografato lato client Amazon S3](#).
7. Scegliere altre opzioni come appropriato per l'applicazione, quindi scegliere Create Cluster (Crea cluster).

Utilizzo di AWS KMS keys per la crittografia EMRFS

La chiave di crittografia AWS KMS deve essere stata creata nella stessa Regione dell'istanza cluster Amazon EMR e dei bucket Amazon S3 utilizzati con EMRFS. Se la chiave specificata si trova in un account diverso da quello utilizzato per configurare un cluster, è necessario specificare la chiave utilizzando il relativo ARN.

Il ruolo del profilo dell'istanza Amazon EC2 deve disporre delle autorizzazioni per utilizzare la chiave KMS specificata. Il ruolo predefinito per il profilo dell'istanza in Amazon EMR è `EMR_EC2_DefaultRole`. Se utilizzi un ruolo diverso per il profilo dell'istanza o utilizzi ruoli IAM per le richieste EMRFS ad Amazon S3, assicurati che ogni ruolo venga aggiunto come utente chiave a seconda dei casi. Ciò concede al ruolo l'autorizzazione a utilizzare la chiave KMS. Per ulteriori informazioni, consulta [Utilizzo di policy delle chiavi](#) nella Guida per gli sviluppatori di AWS Key Management Service e [Configurazione dei ruoli IAM per richieste EMRFS ad Amazon S3](#).

Puoi utilizzare la AWS Management Console per aggiungere il tuo profilo dell'istanza o il profilo dell'istanza EC2 all'elenco degli utenti delle chiavi per la chiave KMS specificata oppure puoi utilizzare la AWS CLI o un SDK AWS per collegare una policy delle chiavi appropriata.

Nota che Amazon EMR supporta solo [Chiavi KMS simmetriche](#). Non è possibile utilizzare una [chiave KMS asimmetrica](#) per crittografare i dati a riposo in un cluster Amazon EMR. Per informazioni su come determinare se una chiave KMS è simmetrica o asimmetrica, consulta [Identificazione di chiavi KMS simmetriche e asimmetriche](#).

La procedura seguente descrive come aggiungere il profilo dell'istanza Amazon EMR predefinito, `EMR_EC2_DefaultRole` come utente di chiavi utilizzando la AWS Management Console.

Presuppone che tu abbia già creato una chiave KMS. Per creare una nuova chiave KMS, consulta [Creazione di chiavi](#) nella Guida per gli sviluppatori di AWS Key Management Service.

Aggiunta del profilo dell'istanza EC2 per Amazon EMR all'elenco degli utenti della chiave di crittografia

1. Accedi alla AWS Management Console e apri la console AWS Key Management Service (AWS KMS) all'indirizzo <https://console.aws.amazon.com/kms>.
2. Per modificare la Regione AWS, utilizza il Selettore di regione nell'angolo in alto a destra della pagina.
3. Seleziona l'alias della chiave KMS da modificare.
4. Nella pagina dei dettagli della chiave, in Key Users (Utenti di chiavi), scegli Add (Aggiungi).
5. Nella finestra di dialogo Add key users (Aggiungi utenti chiave) selezionare il ruolo appropriato. Il nome del ruolo di default è EMR_EC2_DefaultRole.
6. Scegliere Add (Aggiungi).

Crittografia lato server di Amazon S3

Quando configuri la crittografia lato server Amazon S3, Amazon S3 esegue la crittografia dei dati a livello di oggetto, tramite la scrittura dei dati su disco, e ne esegue la decrittografia al momento dell'accesso. Per ulteriori informazioni sulla SEE, consulta [Protezione dei dati con la crittografia lato server](#) nella Guida per l'utente di Amazon Simple Storage Service.

Puoi scegliere tra due diversi sistemi di gestione delle chiavi quando specifichi la SSE in Amazon EMR:

- SSE-S3: Amazon S3 gestisce le chiavi per te.
- SSE-KMS: si utilizza una AWS KMS key configurata con policy adatte per Amazon EMR. Per ulteriori informazioni sui requisiti chiave per Amazon EMR, consulta [Utilizzo di AWS KMS keys per crittografia](#).

La SSE con chiavi fornite dal cliente (SSE-C) non è disponibile per l'utilizzo con Amazon EMR.

Per creare un cluster con SSE-S3 attivato mediante l'AWS CLI

- Digita il seguente comando:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

È anche possibile attivare SSE-S3 impostando la proprietà `fs.s3.enableServerSideEncryption` su `true` nelle proprietà `emrfs-site`. Vedere l'esempio per SSE-KMS esposto di seguito e omettere la proprietà per l'ID di chiave.

Per creare un cluster con SSE-KMS attivato utilizzando l'AWS CLI

Note

SSE-KMS è disponibile solo in Amazon EMR versione 4.5.0 e versioni successive.

- Digitare il seguente comando AWS CLI per creare un cluster con SSE-KMS, dove *keyID* è una AWS KMS key, ad esempio *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]
```

--OPPURE--

Digitare il seguente comando di AWS CLI utilizzando la classificazione `emrfs-site` e fornire un file di configurazione JSON con contenuto simile a `myConfig.json` nell'esempio che segue:

```
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
```

Esempio di contenuto di `myConfig.json`:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.enableServerSideEncryption": "true",
      "fs.s3.serverSideEncryption.kms.keyId": "a4567b8-9900-12ab-1234-123a45678901"
    }
  }
]
```

```

    }
  }
]

```

Proprietà di configurazione per SSE-S3 e SSE-KMS

Queste proprietà possono essere configurate utilizzando la classificazione di configurazione `emrfs-site`. SSE-KMS è disponibile solo in Amazon EMR versione 4.5.0 e versioni successive.

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.enableServerSideEncryption</code>	false	Quando impostata su true , gli oggetti archiviati in Amazon S3 sono crittografati utilizzando la crittografia lato server. Se non viene specificata una chiave, viene utilizzato SSE-S3.
<code>fs.s3.serverSideEncryption.kms.keyId</code>	n/a	Specifica un ARN o un ID di chiave AWS KMS. Se una chiave è specificata, viene utilizzato SSE-KMS.

File crittografato lato client Amazon S3

Con la crittografia lato client di Amazon S3, la crittografia e la decrittografia Amazon S3 avvengono nel client EMRFS nel tuo cluster. Gli oggetti vengono crittografati prima di essere caricati su Amazon S3 e decrittati dopo il loro download. Il provider specificato fornisce la chiave di crittografia utilizzata dal client. Il client può utilizzare le chiavi fornite da AWS KMS (CSE-KMS) o una classe Java personalizzata che fornisce la chiave principale lato client (CSE-C). Le specifiche di crittografia sono leggermente diverse tra CSE-KMS e CSE-C, a seconda del provider specificato e dei metadati dell'oggetto da decrittare o crittografare. Per ulteriori informazioni su queste differenze, consulta [Protezione dei dati tramite la crittografia lato client](#) nella Guida per l'utente di Amazon Simple Storage Service.

Note

Amazon S3 CSE garantisce solo che i dati EMRFS scambiati con Amazon S3 siano crittografati; non tutti i dati sui volumi delle istanze del cluster sono crittografati. Inoltre, poiché Hue non utilizza EMRFS, gli oggetti che il browser di file Hue S3 scrive su Amazon S3 non vengono crittografati.

Specifica di CSE-KMS per i dati di EMRFS in Amazon S3 mediante la AWS CLI

- Digitare il comando seguente e sostituire *MyKMSKeyID* con l'ID di chiave o l'ARN della chiave KMS da utilizzare:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier  
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

Creazione di un provider di chiavi personalizzato

Quando si crea un provider di chiavi personalizzato, l'applicazione dovrebbe implementare l'[interfaccia EncryptionMaterialsProvider](#), che è disponibile in AWS SDK for Java 1.11.0 e versioni successive. L'implementazione può utilizzare qualsiasi strategia per fornire materiali di crittografia. È possibile, ad esempio, scegliere di fornire materiali di crittografia statici o di integrare un sistema di gestione delle chiavi più complesso.

L'algoritmo di crittografia utilizzato per i materiali di crittografia personalizzati deve essere AES/GCM/NoPadding.

La classe EncryptionMaterialsProvider ottiene materiali di crittografia suddivisi in base al contesto. Amazon EMR popola le informazioni sul contesto di crittografia al runtime per aiutare il chiamante a determinare i materiali di crittografia corretti da restituire.

Example Esempio: utilizzo di un provider di chiavi personalizzato per la crittografia Amazon S3 con EMRFS

Quando Amazon EMR recupera i materiali di crittografia dalla classe EncryptionMaterialsProvider per eseguire la crittografia, EMRFS inserisce facoltativamente l'argomento materialsDescription con due campi: l'URI Amazon S3 per l'oggetto e il JobFlowId del cluster, che può essere utilizzato dalla classe EncryptionMaterialsProvider per restituire i materiali di crittografia in modo selettivo.

Ad esempio, il provider potrebbe restituire chiavi diverse per diversi prefissi URI di Amazon S3. Sarà la descrizione dei materiali di crittografia restituiti a essere memorizzata con l'oggetto Amazon S3 anziché il valore `materialsDescription` generato da EMRFS e passato al provider. Durante la decrittografia di un oggetto Amazon S3, la descrizione dei materiali di crittografia viene passata alla classe `EncryptionMaterialsProvider` in modo che possa, ancora una volta, restituire selettivamente la chiave corrispondente per decrittare l'oggetto.

Di seguito viene fornita un'implementazione di riferimento di `EncryptionMaterialsProvider`. Un altro provider personalizzato, [EMRFSRSAEncryptionMaterialsProvider](#), è disponibile da 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;
}
}
```

Configurazione di un provider di materiali personalizzato mediante la AWS CLI

Per utilizzare AWS CLI, passa gli argomenti `Encryption`, `ProviderType`, `CustomProviderClass` e `CustomProviderLocation` all'opzione `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
```

Con l'impostazione di `Encryption` su `ClientSide`, si attiva la crittografia lato client, `CustomProviderClass` è il nome dell'oggetto `EncryptionMaterialsProvider` e `CustomProviderLocation` è la posizione locale o Amazon S3 da cui Amazon EMR copia `CustomProviderClass` in ogni nodo nel cluster e lo inserisce nel classpath.

Configurazione di un provider di materiali personalizzato mediante un kit SDK

Per utilizzare un SDK, puoi impostare la proprietà `fs.s3.cse.encryptionMaterialsProvider.uri` per scaricare la classe `EncryptionMaterialsProvider` personalizzata che memorizzi in Amazon S3 in ogni nodo del cluster. Questa configurazione viene eseguita nel file `emrfs-site.xml` con la CSE abilitata e la corretta posizione del provider personalizzato.

Ad esempio, nel kit AWS SDK for Java che utilizza `RunJobFlowRequest`, il codice potrebbe essere simile a quanto segue:

```

<snip>
    Map<String,String> emrfsProperties = new HashMap<String,String>();
        emrfsProperties.put("fs.s3.cse.encryptedMaterialsProvider.uri","s3://mybucket/
MyCustomEncryptionMaterialsProvider.jar");
        emrfsProperties.put("fs.s3.cse.enabled","true");
        emrfsProperties.put("fs.s3.consistent","true");

    emrfsProperties.put("fs.s3.cse.encryptedMaterialsProvider","full.class.name.of.EncryptionMate

    Configuration myEmrfsConfig = new Configuration()
        .withClassification("emrfs-site")
        .withProperties(emrfsProperties);

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("Custom EncryptionMaterialsProvider")
        .withReleaseLabel("emr-5.36.1")
        .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 personalizzato con argomenti

È possibile che sia necessario passare degli argomenti direttamente al provider. A questo proposito, puoi utilizzare la classificazione di configurazione `emrfs-site` con argomenti personalizzati definiti come proprietà. Un esempio di configurazione, salvato come file `myConfig.json`, è riportato di seguito:

```

[
  {
    "Classification": "emrfs-site",

```

```

    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]

```

Utilizzando il comando `create-cluster` da AWS CLI, puoi utilizzare l'opzione `--configurations` per specificare il file come mostrato di seguito:

```

aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge
--instance-count 2 --configurations file://myConfig.json --emrfs
Encryption=ClientSide,CustomProviderLocation=s3://mybucket/myfolder/
myprovider.jar,CustomProviderClass=classname

```

Configurazione del supporto EMRFS S3EC V2

Per i rilasci S3 Java SDK (1.11.837 e versioni successive), è stato introdotto il client di crittografia versione 2 (S3EC V2) con vari miglioramenti alla sicurezza. Per maggiori informazioni, consulta il post del blog S3 [Updates to the Amazon S3 Encryption Client \(Aggiornamenti al client di crittografia Amazon S3\)](#). Inoltre, fai riferimento alla [Migrazione del client di crittografia Amazon S3](#) nella Guida per gli sviluppatori di AWS SDK for Java.

Il client di crittografia V1 è ancora disponibile nell'SDK per la compatibilità con le versioni precedenti. Per impostazione predefinita, EMRFS utilizzerà S3EC V1 per crittografare e decrittare gli oggetti S3 se CSE è abilitato.

Gli oggetti S3 crittografati con S3EC V2 non possono essere decrittati da EMRFS su un cluster EMR la cui versione di rilascio è precedente a `emr-5.31.0` (`emr-5.30.1` e precedenti, `emr-6.1.0` e precedenti).

Example Configurazione di EMRFS per l'utilizzo di S3EC V2

Per configurare EMRFS per l'utilizzo di S3EC V2, aggiungi la seguente configurazione:

```

{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}

```

}

Proprietà di `emrfs-site.xml` per la crittografia lato client di Amazon S3

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.cse.enabled</code>	false	Quando impostata su true , gli oggetti EMRFS archiviati in Amazon S3 vengono crittografati con la crittografia lato client.
<code>fs.s3.cse.encryptionV2.enabled</code>	false	Se impostato su <code>true</code> , EMRFS utilizza il client di crittografia S3 versione 2 per crittografare e decrittare gli oggetti su S3. Disponibile solo in EMR 5.31.0 e versioni successive.
<code>fs.s3.cse.encryptionMaterialsProvider.uri</code>	N/A	Si applica quando si utilizzano materiali di crittografia personalizzati. L'URI Amazon S3 in cui si trova il file JAR con <code>EncryptionMaterialsProvider</code> . Quando si fornisce questo URI, Amazon EMR scarica automaticamente il JAR in tutti i nodi del cluster.
<code>fs.s3.cse.encryptionMaterialsProvider</code>	N/A	Il percorso della classe <code>EncryptionMaterialsProvider</code> utilizzato con la crittografia lato client. Quando si utilizza CSE-KMS, specificare <code>com.amazon.ws.emr</code> .

Proprietà	Valore predefinito	Descrizione
		<code>hadoop.fs.cse.KMSEncryptionMaterialsProvider</code>
<code>fs.s3.cse.materialsDescription.enabled</code>	false	Quando impostata su <code>true</code> , popola l'argomento <code>materialsDescription</code> degli oggetti crittografati con l'URI di Amazon S3 per l'oggetto e <code>JobFlowId</code> . Da impostare su <code>true</code> se si utilizzano materiali di crittografia personalizzati.
<code>fs.s3.cse.kms.keyId</code>	N/A	Si applica quando si utilizza CSE-KMS. Il valore <code>KeyId</code> , ARN o <code>alias</code> della chiave KMS; utilizzata per la crittografia.
<code>fs.s3.cse.cryptoStorageMode</code>	ObjectMetadata	La modalità di archiviazione Amazon S3. Per impostazione predefinita, la descrizione delle informazioni di crittografia è archiviata nei metadati degli oggetti. È anche possibile archiviare la descrizione in un file di istruzioni. I valori validi sono <code>ObjectMetadata</code> e <code>InstructionFile</code> . Per ulteriori informazioni, consulta Crittografia di dati lato client con la AWS SDK for Java e Amazon S3 .

Delta Lake

Delta Lake è un framework a livello di archiviazione per architetture lakehouse comunemente costruite su Amazon S3. Con Amazon EMR rilascio 6.9.0, puoi utilizzare [Apache Spark 3.x](#) sui cluster Amazon EMR con le tabelle Delta Lake. Per ulteriori informazioni sui lake house con Delta Lake, consulta la pagina <https://delta.io/>.

La seguente tabella indica la versione di Delta inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Delta.

Per la versione dei componenti installati con Delta in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Delta per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Delta	Componenti installati con Delta
emr-6.14.0	Delta 2.4.0	Not available.

Introduzione a Delta Lake

Delta Lake è un progetto open source che aiuta a implementare le moderne architetture di data lake comunemente costruite su Amazon S3. Delta Lake offre le seguenti funzionalità:

- Transazioni atomiche, coerenti, isolate e durevoli (ACID) su Spark. I lettori vedono una visualizzazione coerente della tabella durante un processo Spark.
- Gestione scalabile dei metadati con elaborazione distribuita da Spark.
- Combina casi d'uso in batch e streaming con la stessa tabella Delta.
- Applicazione automatica dello schema per evitare record errati durante l'importazione dei dati.
- Viaggio nel tempo con il controllo delle versioni dei dati.
- Supporta le operazioni di unione, aggiornamento ed eliminazione per casi d'uso complessi come change data capture (CDC), lo streaming di upsert e altro ancora.

Utilizzo di un cluster con Delta Lake installato

Argomenti

- [Utilizzo di un cluster Delta Lake con Flink](#)
- [Utilizzo di un cluster Delta Lake con Trino](#)
- [Utilizzo di un cluster Delta Lake con Spark](#)
- [Utilizzo di un cluster Delta Lake con Spark e AWS Glue](#)

Utilizzo di un cluster Delta Lake con Flink

A partire da Amazon EMR rilascio 6.11, puoi utilizzare Delta Lake con il cluster Flink. Negli esempi seguenti, viene utilizzata la AWS CLI per lavorare con Delta Lake su un cluster Flink di Amazon EMR.

Note

Amazon EMR supporta l'API Flink DataStream quando usi Delta Lake con un cluster Flink.

Creazione di un cluster Delta Lake

1. Creare un file, `delta_configurations.json`, con i seguenti contenuti:

```
[{"Classification":"delta-defaults",  
  "Properties":{"delta.enabled":"true"}}]
```

2. Crea un cluster con la seguente configurazione. Sostituisci il `example` Amazon S3 bucket path e l'`subnet ID` con i valori effettivi.

```
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
```

Inizializzazione di una sessione Flink yarn

Per inizializzare una sessione Flink yarn, esegui il comando riportato di seguito:

```
flink-yarn-session -d
```

Compilazione di un processo Flink con Delta Lake

Gli esempi seguenti mostrano come usare sbt o Maven per compilare il processo Flink con Delta Lake.

sbt

[sbt](#) è uno strumento di compilazione per Scala che puoi usare con poca o nessuna configurazione quando disponi di piccoli progetti.

```
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](#) è uno strumento di automazione della compilazione open-source della Apache Software Foundation. Con Maven, puoi compilare, pubblicare e distribuire un processo Flink con Delta Lake su Amazon EMR.

```
<project>
<properties>
  <scala.main.version>2.12</scala.main.version>
  <delta-connectors-version>0.6.0</delta-connectors-version>
  <flink-version>1.16.1</flink-version>
  <hadoop-version>3.1.0</hadoop-version>
</properties>
```

```
<dependencies>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-flink</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-standalone_${scala-main-version}</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-clients</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-client</artifactId>
    <version>${hadoop-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-common</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-runtime</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
</dependencies>
```

```
</dependencies>
```

Scrittura in una tabella Delta con l'API Flink Datastream

Usa l'esempio seguente per creare un DeltaSink per scrivere nella tabella con un `deltaTablePath`:

```
public static DataStream<RowData> createDeltaSink(  
    DataStream<RowData> stream,  
    String deltaTablePath,  
    RowType rowType) {  
    Configuration configuration = new Configuration();  
    DeltaSink<RowData> deltaSink = DeltaSink  
        .forRowData(  
            new org.apache.flink.core.fs.Path(deltaTablePath),  
            configuration,  
            rowType)  
        .build();  
    stream.sinkTo(deltaSink);  
    return stream;  
}
```

Lettura da una tabella Delta con l'API Flink Datastream

Usa l'esempio seguente per creare un DeltaSource limitato per leggere dalla tabella con un `deltaTablePath`:

```
public static DataStream<RowData> createBoundedDeltaSourceAllColumns(  
    StreamExecutionEnvironment env,  
    String deltaTablePath) {  
    Configuration configuration = new Configuration();  
    DeltaSource<RowData> deltaSource = DeltaSource  
        .forBoundedRowData(  
            new org.apache.flink.core.fs.Path(deltaTablePath),  
            configuration)  
        .build();  
  
    return env.fromSource(deltaSource, WatermarkStrategy.noWatermarks(), "delta-  
source");  
}
```

Creazione di sink con supporto multi-cluster per Delta Lake standalone

Usa l'esempio seguente per creare un DeltaSink per scrivere su una tabella con un `deltaTablePath` e un supporto [multi-cluster](#):

```
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;  
}
```

Esecuzione del processo Flink

Per eseguire il processo, utilizza il comando seguente:

```
flink run FlinkJob.jar
```

Utilizzo di un cluster Delta Lake con Trino

Con Amazon EMR rilascio 6.9.0 e successivi, puoi utilizzare Delta Lake con il cluster Trino.

In questo tutorial, useremo la AWS CLI per lavorare con Delta Lake su un cluster Amazon EMR Trino.

Creazione di un cluster Delta Lake

1. Crea un file, `delta_configurations.json`, e imposta i valori per il catalogo scelto. Ad esempio, se desideri utilizzare il metastore Hive come catalogo, il file deve avere il seguente contenuto:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

Se desideri utilizzare AWS Glue Catalog come archivio, il file deve avere il seguente contenuto:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. Crea un cluster con la seguente configurazione, sostituendo **example Amazon S3 bucket path** e **subnet ID** con i tuoi valori.

```
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
```

Inizializzazione della sessione Trino per Delta Lake

Per inizializzare una sessione Trino, esegui il comando seguente

```
trino-cli --catalog delta
```

Scrittura in una tabella Delta Lake

Crea e scrivi nella tabella utilizzando i comandi SQL seguenti:

```
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');
```

Lettura da una tabella Delta Lake

Leggi dalla tabella con il comando SQL seguente:

```
SELECT * from default.delta_table;
```

Utilizzo di un cluster Delta Lake con Spark

A partire da Amazon EMR versione 6.9.0, puoi utilizzare Delta Lake con il cluster Spark senza dover eseguire operazioni di bootstrap. Per Amazon EMR rilascio 6.8.0 e precedenti, puoi utilizzare le operazioni di bootstrap per preinstallare le dipendenze necessarie.

Negli esempi seguenti, viene utilizzata la AWS CLI per lavorare con Delta Lake su un cluster Spark di Amazon EMR.

Per utilizzare Delta Lake su Amazon EMR con la AWS Command Line Interface, crea prima un cluster. Per informazioni su come specificare la classificazione Delta Lake con la AWS Command Line Interface, consulta la sezione [Fornitura di una configurazione nella AWS Command Line Interface quando si crea un cluster](#) o [Fornitura di una configurazione con l'SDK Java quando si crea un cluster](#).

1. Creare un file, `configurations.json`, con i seguenti contenuti:

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} ]]
```

2. Crea un cluster con la seguente configurazione, sostituendo **bucket path** e **subnet ID** di Amazon S3 di esempio con i tuoi valori.

```
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
```

In alternativa, puoi creare un cluster Amazon EMR e l'applicazione Spark con i file seguenti come dipendenze JAR in un processo Spark:

```
/usr/share/aws/delta/lib/delta-core.jar,
/usr/share/aws/delta/lib/delta-storage.jar,
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar
```

Per ulteriori informazioni, consulta [Invio di applicazioni](#).

Per includere una dipendenza jar in un processo Spark, è possibile aggiungere le seguenti proprietà di configurazione all'applicazione Spark:

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-core.jar,
/usr/share/aws/delta/lib/delta-storage.jar,
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

Per ulteriori informazioni sulle dipendenze dei processi Spark, consulta la sezione [Dependency Management](#) (Gestione delle dipendenze).

Inizializzazione della sessione Spark per Delta Lake

Negli esempi seguenti viene illustrato come avviare la shell (interprete di comandi) interattiva Spark, utilizzare Spark-submit o utilizzare i notebook Amazon EMR per lavorare con Delta Lake su Amazon EMR.

spark-shell

1. Effettua la connessione al nodo primario tramite SSH. Per ulteriori informazioni, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida alla gestione di Amazon EMR.
2. Immettere il seguente comando per avviare la shell Spark. Per utilizzare la shell PySpark, sostituire `spark-shell` con `pyspark`.

```
spark-shell \  
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \  
  --conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

spark-submit

1. Effettua la connessione al nodo primario tramite SSH. Per ulteriori informazioni, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida alla gestione di Amazon EMR.
2. Inserisci il seguente comando per avviare la sessione Spark per Delta Lake.

```
spark-submit  
  -conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"  
  -conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

EMR Studio notebooks

Per inizializzare una sessione Spark utilizzando i notebook Amazon EMR Studio, configura la sessione Spark utilizzando il comando `%%configure magic` nel notebook Amazon EMR, come nell'esempio seguente. Per ulteriori informazioni, consulta [Utilizzo di EMR Notebooks magics](#) nella Guida alla gestione di Amazon EMR.

```
%%configure -f  
{  
  "conf": {  
    "spark.sql.extensions": "io.delta.sql.DeltaSparkSessionExtension",  
    "spark.sql.catalog.spark_catalog":  
    "org.apache.spark.sql.delta.catalog.DeltaCatalog"
```



```
}  
}
```

Scrittura in una tabella Delta Lake

Nell'esempio seguente viene illustrato come creare un DataFrame e scriverlo come set di dati Delta Lake. In questa sezione, gli esempi illustrano l'utilizzo di set di dati con la shell (interprete di comandi) Spark connessi al nodo primario utilizzando SSH come utente hadoop predefinito.

Note

Per incollare gli esempi di codice nella shell Spark, digitare `:paste` al prompt, incollare l'esempio e premere CTRL + D.

PySpark

Spark include anche una shell (interprete di comandi) basata su Python, `pyspark`, che puoi utilizzare per realizzare prototipi di programmi Spark scritti in Python. Proprio come con `spark-shell`, invoca `pyspark` sul nodo primario.

```
## 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");
```

Lettura da una tabella Delta Lake

PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

Scala

```
val ddf = spark.table("delta_table")
ddf.show()
```

SQL

```
SELECT * FROM delta_table;
```

Utilizzo di un cluster Delta Lake con Spark e AWS Glue

Per utilizzare AWS Glue Catalog come metastore per le tabelle Delta Lake, crea un cluster con i passaggi seguenti. Per informazioni su come specificare la classificazione Delta Lake utilizzando la AWS Command Line Interface, consulta la sezione [Fornitura di una configurazione nella AWS Command Line Interface quando si crea un cluster](#) o [Fornitura di una configurazione utilizzando l'SDK Java quando si crea un cluster](#).

Creazione di un cluster Delta Lake

1. Creare un file, `configurations.json`, con i seguenti contenuti:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"spark-hive-site",
  "Properties":
  {"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Crea un cluster con la seguente configurazione, sostituendo **example Amazon S3 bucket path** e **subnet ID** con i tuoi valori.

```
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
```

Considerazioni e limitazioni

- Delta Lake è supportato per l'uso con Amazon EMR rilasci 6.9.0 e successivi. Puoi usare [Apache Spark 3.x](#) su cluster Amazon EMR con tabelle Delta.
- Ti consigliamo di utilizzare lo schema URI s3 per i percorsi di posizione S3 anziché s3a per garantire i massimi livelli di prestazioni, sicurezza e affidabilità. Per ulteriori informazioni, consulta la sezione [Working with storage and file systems](#) (Utilizzo dell'archiviazione e dei file system).
- Con Amazon EMR 6.9 e 6.10, quando i dati delle tabelle Delta Lake vengono archiviati in Amazon S3, i dati delle colonne diventano NULL dopo l'operazione di ridenominazione delle colonne. Questo problema è stato risolto con Amazon EMR 6.11. Per ulteriori informazioni sull'operazione sperimentale di ridenominazione delle colonne, consulta [Operazione di ridenominazione delle colonne](#) nella Guida per l'utente di Delta Lake.
- Se crei un database nel Catalogo dati AWS Glue al di fuori di Apache Spark, il database potrebbe avere un campo LOCATION vuoto. Poiché Spark non consente di creare database con una proprietà di posizione vuota, si otterrà il seguente errore se si utilizza Spark in Amazon EMR per creare una tabella Delta in un database Glue e il database ha una proprietà LOCATION vuota:

```
IllegalArgumentException: Can not create a Path from an empty string
```

Per risolvere questo problema, crea il database nel Catalogo dati con un percorso valido e non vuoto per il campo LOCATION. Per i passaggi da seguire per implementare questa soluzione, consulta [Eccezione di argomento non valido durante la creazione di una tabella](#) nella Guida per l'utente di Amazon Athena.

Cronologia dei rilasci Delta

Nella seguente tabella sono elencati la versione di Delta inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione Versione componente per il rilascio in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Delta

Etichetta di rilascio di Amazon EMR	Versione di Delta	Componenti installati con Delta
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](#) è un motore di flussi di dati in streaming che può essere utilizzato per l'elaborazione di flussi in tempo reale su origini dati a elevato throughput. Flink supporta semantica di eventi temporali per eventi di esaurimento, semantica di elaborazione singola, controlli di congestione e API ottimizzate per la scrittura con applicazioni in streaming e in batch.

Inoltre, Flink dispone di connettori per origini dati di terze parti, ad esempio le seguenti:

- [Flusso di dati Amazon Kinesis](#)
- [Apache Kafka](#)
- [Connettore Flink Elasticsearch](#)
- [API di streaming di Twitter](#)
- [Cassandra](#)

Amazon EMR supporta Flink come applicazione YARN in modo da poter gestire le risorse insieme ad altre applicazioni all'interno di un cluster. Flink-on-YARN consente di inviare lavori Flink transitori oppure di creare un cluster di lunga durata che accetta più lavori e alloca le risorse in base alla prenotazione YARN complessiva.

Flink è incluso in Amazon EMR 5.1.0 e versioni successive.

Note

Nella versione 5.2.1 di Amazon EMR è stato integrato il supporto per la classe `FlinkKinesisConsumer`.

La seguente tabella indica la versione di Flink inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Flink.

Per la versione dei componenti installati con Flink in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Flink per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-6.14.0	Flink 1.17.1-amzn-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-resource manager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

La seguente tabella indica la versione di Flink inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Flink.

Per la versione dei componenti installati con Flink in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Flink per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-5.36.1	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-resource manager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Argomenti

- [Creazione di un cluster con Flink](#)
- [Configurazione di Flink in Amazon EMR](#)
- [Utilizzo dei processi Flink in Amazon EMR](#)
- [Utilizzo della shell Scala](#)
- [Trovare l'interfaccia Web Flink](#)
- [Utilizzo dei processi Flink di Zeppelin in Amazon EMR](#)
- [Cronologia dei rilasci di Flink](#)

Creazione di un cluster con Flink

Puoi avviare un cluster con la AWS Management Console, la AWS CLI o un SDK AWS.

Per avviare un cluster con Flink installato dalla console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Per Software Configuration (Configurazione software), scegli EMR Release emr-5.1.0 o successivo.
4. Scegli Flink come applicazione, insieme alle altre da installare.
5. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Per avviare un cluster con Flink dalla AWS CLI

- Creare il cluster con il seguente comando:

```
aws emr create-cluster --release-label emr-5.36.1 \  
--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

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

Configurazione di Flink in Amazon EMR

Configurazione di Flink con Hive Metastore e Catalogo Glue

Amazon EMR rilascio 6.9.0 e successivi supportano sia Hive Metastore sia Catalogo AWS Glue con il connettore Apache Flink per Hive. Questa sezione descrive i passaggi necessari per configurare [Catalogo AWS Glue](#) e [Hive Metastore](#) con Flink.

Argomenti

- [Utilizzo di Hive Metastore](#)
- [Utilizza il Catalogo dati AWS Glue](#)

Utilizzo di Hive Metastore

1. Crea un cluster EMR con il rilascio 6.9.0 o successivi e almeno due applicazioni: Hive e FLINK.
2. Utilizza [script runner](#) per eseguire il seguente script come funzione passo passo:

```
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

Utilizza il Catalogo dati AWS Glue

1. Crea un cluster EMR con il rilascio 6.9.0 o successivi e almeno due applicazioni: Hive e FLINK.
2. Seleziona Utilizza per i metadati delle tabelle Hive nelle impostazioni del Catalogo dati AWS Glue per abilitare il Catalogo dati nel cluster.
3. Utilizza [script runner](#) per eseguire il seguente script come funzione di fase: [Esegui comandi e script su un cluster Amazon EMR](#):

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

Configurazione di Flink con un file di configurazione

Puoi utilizzare l'API di configurazione di Amazon EMR per configurare Flink con un file di configurazione. I file configurabili all'interno dell'API sono:

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

Il file di configurazione principale per Flink è `flink-conf.yaml`.

Per configurare il numero di slot di attività utilizzati per Flink dalla AWS CLI

1. Creare un file, `configurations.json`, con i seguenti contenuti:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

]

2. Quindi, creare un cluster con la seguente configurazione:

```
aws emr create-cluster --release-label emr-5.36.1 \  
--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

Puoi anche modificare alcune configurazioni con l'API Flink. Per ulteriori informazioni, consulta [Concepts \(Concetti\)](#) nella documentazione di Flink.

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Opzioni di parallelismo

Il proprietario dell'applicazione sa quali risorse assegnare alle attività all'interno di Flink. Per gli esempi riportati in questa documentazione, utilizza lo stesso numero di attività delle istanze di attività utilizzate per l'applicazione. Generalmente lo consigliamo per il livello iniziale di parallelismo, ma è anche possibile aumentare la granularità del parallelismo con gli slot delle attività, che generalmente non dovrebbero superare il numero di [core virtuali](#) per istanza. Per ulteriori informazioni sull'architettura Flink, consulta [Concetti](#) nella documentazione di Flink.

Configurazione di Flink su un cluster EMR con nodi primari multipli

Il JobManager di Flink rimane disponibile durante il processo di failover del nodo primario in un cluster Amazon EMR con più nodi primari. A partire da Amazon EMR 5.28.0, anche la disponibilità

elevata di JobManager è abilitata automaticamente. Non è necessaria alcuna configurazione manuale.

Con le versioni 5.27.0 o precedenti di Amazon EMR, JobManager è un singolo punto di errore. Quando il JobManager ha esito negativo, perde tutti gli stati del processo e non riprende i processi in esecuzione. Puoi abilitare l'elevata disponibilità di JobManager configurando il numero di tentativi dell'applicazione, il checkpoint e l'abilitazione di ZooKeeper come storage di stato per Flink, come dimostra l'esempio seguente:

```
[
  {
    "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"
    }
  }
]
```

È necessario configurare il numero massimo di tentativi master dell'applicazione per YARN e i tentativi dell'applicazione per Flink. Per ulteriori informazioni, consulta [Configurazione dell'elevata disponibilità del cluster YARN](#). È anche possibile configurare il checkpoint di Flink per fare in modo che JobManager riavviato ripristini i processi in esecuzione da checkpoint completati in precedenza. Per ulteriori informazioni, consulta [Checkpoint di Flink](#).

Configurazione delle dimensioni del processo di memoria

Per le versioni Amazon EMR che utilizzano Flink 1.11.x, è necessario configurare la dimensione totale del processo di memoria sia per JobManager (`jobmanager.memory.process.size`) che per TaskManager (`taskmanager.memory.process.size`) in `flink-conf.yaml`. È possibile impostare questi valori configurando il cluster con l'API di configurazione o rimuovendo manualmente i commenti da questi campi tramite SSH. Flink fornisce i seguenti valori di default.

- `jobmanager.memory.process.size`: 1600m
- `taskmanager.memory.process.size`: 1728m

Per escludere il metaspace e l'overhead JVM, utilizza la dimensione totale della memoria Flink (`taskmanager.memory.flink.size`) anziché `taskmanager.memory.process.size`. Il valore di default per `taskmanager.memory.process.size` è 1280m. Non è consigliabile impostare sia `taskmanager.memory.process.size` che `taskmanager.memory.flink.size`.

Tutte le versioni di Amazon EMR che utilizzano Flink 1.12.0 e versioni successive hanno i valori di default elencati nell'open-source impostati per Flink come valori predefiniti su Amazon EMR, quindi non è necessario configurarli personalmente.

Configurazione della dimensione del file di output del log

I container di applicazioni Flink creano e scrivono in tre tipi di file di log: file `.out`, file `.log` e file `.err`. Solo i file `.err` vengono compressi e rimossi dal file system, mentre i file di log `.log` e `.out` rimangono nel file system. Per garantire che questi file di output rimangano gestibili e che il cluster rimanga stabile, è possibile configurare la rotazione dei log in `log4j.properties` per impostare un numero massimo di file e limitarne le dimensioni.

Amazon EMR versione 5.30.0 e successive

A partire dalla versione 5.30.0 di Amazon EMR, Flink utilizza il framework di registrazione `log4j2` con il nome della classificazione di configurazione `flink-log4j`. Nella seguente configurazione di esempio viene illustrato il formato `log4j2`.

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "appender.main.name": "MainAppender",
      "appender.main.type": "RollingFile",
      "appender.main.append" : "false",
      "appender.main.fileName" : "${sys:log.file}",
      "appender.main.filePattern" : "${sys:log.file}.%i",
      "appender.main.layout.type" : "PatternLayout",
      "appender.main.layout.pattern" : "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x - %m
%n",
      "appender.main.policies.type" : "Policies",
      "appender.main.policies.size.type" : "SizeBasedTriggeringPolicy",
```

```

    "appender.main.policies.size.size" : "100MB",
    "appender.main.strategy.type" : "DefaultRolloverStrategy",
    "appender.main.strategy.max" : "10"
  },
}
]

```

Amazon EMR versione 5.29.0 e precedenti

Con Amazon EMR versioni 5.29.0 e precedenti, Flink utilizza il framework di registrazione log4j. La seguente configurazione di esempio illustra il formato log4j

```

[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "log4j.appender.file": "org.apache.log4j.RollingFileAppender",
      "log4j.appender.file.append":"true",
      # keep up to 4 files and each file size is limited to 100MB
      "log4j.appender.file.MaxFileSize":"100MB",
      "log4j.appender.file.MaxBackupIndex":4,
      "log4j.appender.file.layout":"org.apache.log4j.PatternLayout",
      "log4j.appender.file.layout.ConversionPattern":"%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p
%-60c %x - %m%n"
    },
  }
]

```

Configurazione di Flink per l'esecuzione con Java 11

I rilasci 6.12.0 e successivi di Amazon EMR forniscono il supporto di runtime Java 11 per Flink. Le sezioni seguenti descrivono come configurare il cluster per fornire il supporto di runtime Java 11 per Flink.

Argomenti

- [Configurazione di Flink per Java 11 quando si crea un cluster](#)
- [Configurazione di Flink per Java 11 su un cluster in esecuzione](#)
- [Verifica del runtime Java per Flink su un cluster in esecuzione](#)

Configurazione di Flink per Java 11 quando si crea un cluster

Per creare un cluster EMR con Flink e il runtime Java 11, procedi come segue. Il file di configurazione in cui aggiungere il supporto per il runtime Java 11 è `flink-conf.yaml`.

New console

Per creare un cluster con Flink e il runtime Java 11 nella nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegli Cluster in EMR su EC2 nel riquadro di navigazione, quindi scegli Crea cluster.
3. Seleziona Amazon EMR rilascio 6.12.0 o successivi e scegli di installare l'applicazione Flink. Seleziona tutte le altre applicazioni che desideri installare sul tuo cluster.
4. Continua a configurare il cluster. Nella sezione Impostazioni software opzionali, utilizza l'opzione predefinita Inserisci configurazione e inserisci la seguente configurazione:

```
[
  {
    "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. Continua a configurare e avvia il cluster.

AWS CLI

Per creare un cluster con Flink e il runtime Java 11 dalla CLI

1. Crea un file di configurazione `configurations.json` che configuri Flink per l'utilizzo di Java 11.

```
[
  {
    "Classification": "flink-conf",
```



```
"Properties": {
  "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
  "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
  "env.java.home":"/usr/lib/jvm/jre-11"
}
}
```

2. Dalla AWS CLI, crea un nuovo cluster EMR con Amazon EMR rilascio 6.12.0 o successivi e installa l'applicazione Flink, come mostrato nell'esempio seguente:

```
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
```

Configurazione di Flink per Java 11 su un cluster in esecuzione

Utilizza i seguenti passaggi per aggiornare un cluster EMR in esecuzione con Flink e il runtime Java 11. Il file di configurazione in cui aggiungere il supporto per il runtime Java 11 è `flink-conf.yaml`.

New console

Per aggiornare un cluster in esecuzione con Flink e il runtime Java 11 nella nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegli Cluster in EMR su EC2 nel riquadro di navigazione e seleziona il cluster da aggiornare.

Note

Il cluster deve utilizzare Amazon EMR rilascio 6.12.0 o successivi per supportare Java 11.

3. Seleziona la scheda Configurazione.
4. Nella sezione Configurazioni del gruppo di istanze, seleziona il gruppo di istanze In esecuzione che desideri aggiornare, quindi scegli Riconfigura dal menu delle azioni dell'elenco.
5. Riconfigura il gruppo di istanze con l'opzione Modifica attributi come segue. Seleziona Aggiungi nuova configurazione dopo ognuna di esse.

Classificazione	Proprietà	Valore
flink-conf	containerized.task manager.env.JAVA_H OME	/usr/lib/jvm/jre-1 1
flink-conf	containerized.mast er.env.JAVA_HOME	/usr/lib/jvm/jre-1 1
flink-conf	env.java.home	/usr/lib/jvm/jre-1 1

6. Quindi seleziona Salva modifiche per aggiungere le configurazioni.

AWS CLI

Per aggiornare un cluster in esecuzione per utilizzare Flink e il runtime Java 11 dalla CLI

Utilizza il comando `modify-instance-groups` per specificare una nuova configurazione per un gruppo di istanze in un cluster in esecuzione.

1. Innanzitutto, crea un file di configurazione `configurations.json` che configuri Flink per l'utilizzo di Java 11. Nell'esempio seguente, sostituisci `ig-1xxxxxxxx9` con l'ID del gruppo di istanze che desideri riconfigurare. Salva il file nella stessa directory in cui eseguirai il comando `modify-instance-groups`.

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxxx9",
    "Configurations": [
      {
        "Classification": "flink-conf",
```

```

        "Properties":{
            "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
            "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
            "env.java.home":"/usr/lib/jvm/jre-11"
        },
        "Configurations":[]
    }
]
}
]

```

2. Dalla AWS CLI, utilizza il comando seguente. Sostituisci l'ID per il gruppo di istanze che desideri riconfigurare:

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

Verifica del runtime Java per Flink su un cluster in esecuzione

Per determinare il runtime Java per un cluster in esecuzione, accedi al nodo primario con SSH come descritto in [Connessione al nodo primario con SSH](#). Quindi, esegui il comando riportato di seguito:

```
ps -ef | grep flink
```

Il comando `ps` con l'opzione `-ef` elenca tutti i processi in esecuzione sul sistema. È possibile filtrare l'output con `grep` per trovare le menzioni della stringa `flink`. Esamina l'output per il valore Java Runtime Environment (JRE), `jre-XX`. Nell'output seguente, `jre-11` indica che Java 11 viene rilevato in fase di runtime per Flink.

```
flink    19130    1  0 09:17 ?        00:00:15 /usr/lib/jvm/jre-11/bin/
java -Djava.io.tmpdir=/mnt/tmp -Dlog.file=/usr/lib/flink/log/flink-flink-
historyserver-0-ip-172-31-32-127.log -Dlog4j.configuration=file:/usr/lib/flink/conf/
log4j.properties -Dlog4j.configurationFile=file:/usr/lib/flink/conf/log4j.properties
-Dlogback.configurationFile=file:/usr/lib/flink/conf/logback.xml -classpath /usr/lib/
flink/lib/flink-cep-1.17.0.jar:/usr/lib/flink/lib/flink-connector-files-1.17.0.jar:/
usr/lib/flink/lib/flink-csv-1.17.0.jar:/usr/lib/flink/lib/flink-json-1.17.0.jar:/usr/
lib/flink/lib/flink-scala_2.12-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-java-
uber-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-scala-bridge_2.12-1.17.0.
```

In alternativa, [accedi al nodo primario con SSH](#) e avvia una sessione Flink YARN con il comando `flink-yarn-session -d`. L'output mostra la Java Virtual Machine (JVM) per Flink, `java-11-amazon-corretto` nel seguente esempio:

```
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
```

Utilizzo dei processi Flink in Amazon EMR

Esistono diversi modi per interagire con Flink su Amazon EMR: tramite la console, tramite l'interfaccia Flink che si trova sull'interfaccia utente di ResourceManager Tracking e sulla riga di comando. Puoi inviare un file JAR a un'applicazione Flink in uno di questi modi. Una volta inviato un file JAR, il file diventa un processo gestito dal JobManager Flink. Il JobManager si trova sul nodo YARN che ospita il daemon Application Master della sessione Flink.

Puoi eseguire un'applicazione Flink come processo YARN su un cluster di lunga durata o su un cluster transitorio. Su un cluster di lunga durata, è possibile inviare più processi Flink a un cluster Flink in esecuzione su Amazon EMR. Se si esegue un processo Flink su un cluster transitorio, il cluster Amazon EMR esiste solo per il tempo necessario all'esecuzione dell'applicazione Flink, quindi verranno addebitati solo le risorse e il tempo utilizzati. In entrambi i casi, è possibile inviare un processo Flink con l'operazione API `AddSteps` di Amazon EMR, come argomento di fase all'operazione `RunJobFlow` e tramite i comandi della AWS CLI `add-steps` o `create-cluster`.

Avvio di un'applicazione YARN di Flink come fase su un cluster di lunga durata

Per avviare un'applicazione Flink alla quale più client possono inviare lavoro tramite operazioni API YARN, è necessario creare un cluster o aggiungere a un'applicazione Flink un cluster esistente. Per istruzioni su come creare un nuovo cluster, consulta [Creazione di un cluster con Flink](#). Per avviare una sessione YARN su un cluster esistente, segui le seguenti fasi dalla console, dalla AWS CLI o dall'SDK Java.

Note

Il comando `flink-yarn-session` è stato aggiunto alla versione 5.5.0 di Amazon EMR come wrapper per lo script `yarn-session.sh` per semplificare l'esecuzione. Se usi una

versione precedente di Amazon EMR, sostituisci `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` per Arguments (Argomenti) nella console o Args nel comando AWS CLI.

Per inviare un processo Flink su un cluster esistente dalla console

Invia la sessione Flink usando il comando `flink-yarn-session` in un cluster esistente.

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Nell'elenco dei cluster, selezionare il cluster precedentemente avviato.
3. Nella pagina dei dettagli del cluster, scegliere Steps (Fasi), Add Step (Aggiungi fase).
4. Utilizza le linee guida che seguono per immettere i parametri, quindi scegli Aggiungi.

Parametro	Descrizione
Step type (Tipo di fase)	Custom JAR (JAR personalizzato)
Nome	Un nome che aiuta a identificare la nuova fase. Ad esempio, <code><example-flink-step-name></code> .
Jar location (Ubicazione del Jar)	command-runner.jar
Arguments (Argomenti)	Il comando <code>flink-yarn-session</code> con argomenti appropriati per l'applicazione. Ad esempio, <code>flink-yarn-session -d</code> avvia una sessione Flink all'interno del cluster YARN in uno stato scollegato (-d). Consulta Impostazione YARN nella documentazione più aggiornata di Flink per i dettagli sugli argomenti.

Per inviare un processo Flink su un cluster esistente con la AWS CLI

- Utilizza il comando `add-steps` per avviare un processo Flink all'interno di un cluster di lunga durata. Il seguente comando di esempio specifica `Args="flink-yarn-session", "-d"` per

avviare una sessione Flink all'interno del cluster YARN in uno stato scollegato (-d). Consulta [Impostazione YARN](#) nella documentazione più aggiornata di Flink per i dettagli sugli argomenti.

```
aws emr add-steps --cluster-id <j-XXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-fLink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

Invio di lavoro a un'applicazione Flink esistente su un cluster di lunga durata

Se si dispone già di un'applicazione Flink esistente su un cluster di lunga durata, è possibile specificare l'ID applicazione Flink del cluster per inviare il lavoro. Per ottenere l'ID applicazione, esegui `yarn application -list` sulla AWS CLI o tramite l'operazione API [YarnClient](#):

```
$ yarn application -list
16/09/07 19:32:13 INFO client.RMProxy: Connecting to ResourceManager at
ip-10-181-83-19.ec2.internal/10.181.83.19:8032
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED,
RUNNING]):1
Application-Id      Application-Name      Application-Type      User      Queue      State
Final-State      Progress      Tracking-URL
application_1473169569237_0002      Flink session with 14 TaskManagers (detached)
Apache Flink      hadoop      default      RUNNING      UNDEFINED
100% http://ip-10-136-154-194.ec2.internal:33089
```

L'ID applicazione per questa sessione Flink è `application_1473169569237_0002`, che è possibile utilizzare per inviare il lavoro all'applicazione utilizzando la AWS CLI o un SDK.

Example SDK per 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>
```

Invio di un processo transitorio Flink

I seguenti esempi avviano un cluster transitorio che esegue un processo Flink e poi termina al completamento.

Example SDK per Java

```
import java.util.ArrayList;
import java.util.List;
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;

public class Main_test {

    public static void main(String[] args) {
        AWSCredentials credentials_profile = null;
        try {
            credentials_profile = new ProfileCredentialsProvider("default").getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                "Cannot load credentials from .aws/credentials file. " +

```

```
        "Make sure that the credentials file exists and the profile name is
specified within it.",
        e);
    }

AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
    .withRegion(Regions.US_WEST_1)
    .build();

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();
HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash", "-c", "flink", "run", "-m", "yarn-cluster", "-yn", "2", "/usr/
lib/flink/examples/streaming/WordCount.jar", "--input", "s3://path/to/input-file.txt",
"--output", "s3://path/to/output/");

StepConfig flinkRunWordCountStep = new StepConfig()
    .withName("Flink add a wordcount step and terminate")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCountStep);

Application flink = new Application().withName("Flink");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("flink-transient")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(flink)
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withLogUri("s3://path/to/my/logfiles")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myEc2Key")
        .withEc2SubnetId("subnet-12ab3c45")
        .withInstanceCount(3)
        .withKeepJobFlowAliveWhenNoSteps(false)
        .withMasterInstanceType("m4.large")
        .withSlaveInstanceType("m4.large"))
    .withSteps(stepConfigs);

RunJobFlowResult result = emr.runJobFlow(request);
System.out.println("The cluster ID is " + result.toString());
```



```
}  
  
}
```

Example AWS CLI

Utilizza il sottocomando `create-cluster` per creare un cluster transitorio che termina al termine del processo Flink:

```
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/""
```

Utilizzo della shell Scala

Attualmente, la shell Scala di Flink per i cluster EMR è configurata solo per avviare nuove sessioni YARN. È possibile utilizzare la shell Scala seguendo la procedura riportata di seguito.

Utilizzo della shell Scala di Flink sul nodo primario

1. Accedi al nodo primario con SSH come descritto in [Connessione al nodo primario con SSH](#).
2. Per avviare una shell, digitare quanto segue:

In Amazon EMR versione 5.5.0 e successive, è possibile utilizzare il seguente comando per avviare un cluster Yarn per la shell Scala con un TaskManager.

```
% flink-scala-shell yarn 1
```

Nelle versioni precedenti di Amazon EMR, utilizza:

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```


In questo modo si avvia la shell di Flink Scala in modo da poter utilizzare Flink interattivamente. Come per altre interfacce e opzioni, è possibile dimensionare il valore dell'opzione `-n` utilizzato nell'esempio in base al numero di attività che si desidera eseguire dalla shell.

Per ulteriori informazioni, consulta [Scala REPL](#) nella documentazione ufficiale di Apache Flink.

Trovare l'interfaccia Web Flink

L'Application Master che appartiene all'applicazione Flink ospita l'interfaccia Web Flink. È un modo alternativo per inviare un JAR come processo o per visualizzare lo stato corrente di altri processi. L'interfaccia web di Flink è attiva finché è in corso una sessione Flink. Se è già attivo un processo YARN di lunga durata, è possibile seguire le istruzioni dell'argomento [Connessione al nodo primario con SSH](#) nella Guida alla gestione di Amazon EMR per connettersi al ResourceManager YARN. Per esempio, se hai configurato un tunnel SSH e attivato un proxy nel browser, scegli la connessione ResourceManager in Connections (Connessioni) nella pagina dei dettagli del cluster EMR.

Cluster: Development Cluster Waiting Cluster ready after last step completed.

Connections:  [Resource Manager ... \(View All\)](#)

Dopo aver trovato ResourceManager, seleziona l'applicazione YARN che ospita una sessione di Flink. Scegliere il link sotto la colonna Tracking UI (Monitoraggio IU).

All Applications

Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealed Nodes
2	2 GB	11.25 GB	0 B	2	8	0	1	0	0	0	0

Scheduling Resource Type	Minimum Allocation	Maximum Allocation
MEMORY	<memory:32, vCores:1>	<memory:11520, vCores:8>

Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
Flink session with 1 TaskManagers (detached)	Apache Flink	default	Mon Oct 10 14:42:47 -0700 2016	N/A	RUNNING	UNDEFINED		ApplicationMaster

First Previous

Nell'interfaccia Web di Flink, è possibile visualizzare la configurazione, inviare il proprio JAR personalizzato come processo o monitorare i processi in corso.

The screenshot shows the Apache Flink Dashboard with the following components:

- Overview:** Version: 1.1.1
- Task Managers:** 1 Task Managers
- Task Slots:** 1 Task Slots
- Available Task Slots:** 1 Available Task Slots
- Total Jobs:**
 - Running: 0
 - Finished: 0
 - Canceled: 0
 - Failed: 0
- Running Jobs:** Table with columns: Start Time, End Time, Duration, Job Name, Job ID, Tasks, Status.
- Completed Jobs:** Table with columns: Start Time, End Time, Duration, Job Name, Job ID, Tasks, Status.

Utilizzo dei processi Flink di Zeppelin in Amazon EMR

Introduzione

I rilasci 6.10.0 e successivi di Amazon EMR supportano l'integrazione [Apache Zeppelin](#) con Apache Flink. Puoi inviare processi Flink in modo interattivo tramite i notebook Zeppelin. Con l'interprete

Flink, puoi eseguire query Flink, definire lo streaming Flink e i processi in batch e visualizzare l'output all'interno dei notebook Zeppelin. L'interprete Flink è costruito sulla base della REST API di Flink. Ciò consente di accedere e manipolare i processi Flink dall'ambiente Zeppelin per eseguire l'elaborazione e l'analisi dei dati in tempo reale.

Ci sono quattro sottointerpreti nell'interprete Flink. Servono a scopi diversi, ma si trovano tutti nella JVM e condividono gli stessi punti di ingresso preconfigurati per Flink (`ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). Gli interpreti sono i seguenti:

- `%flink` – Crea `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment` `StreamTableEnvironment` e fornisce un ambiente Scala
- `%flink.pyflink` – Fornisce un ambiente Python
- `%flink.ssql` – Fornisce un ambiente SQL in streaming
- `%flink.bsql` – Fornisce un ambiente SQL in batch

Prerequisiti

- L'integrazione di Zeppelin con Flink è supportata per i cluster creati con Amazon EMR 6.10.0 e versioni successive.
- Per visualizzare le interfacce Web ospitate sui cluster EMR come richiesto per questi passaggi, è necessario configurare un tunnel SSH per consentire l'accesso in entrata. Per ulteriori informazioni, consulta la sezione [Configurazione delle impostazioni proxy per visualizzare siti Web ospitati sul nodo primario](#).

Configurazione di Zeppelin-Flink su un cluster EMR

Utilizza i seguenti passaggi per configurare Apache Flink su Apache Zeppelin per l'esecuzione su un cluster EMR:


1. Crea un nuovo cluster dalla console Amazon EMR. Seleziona `emr-6.10.0` o versioni successive per il rilascio di Amazon EMR. Quindi, scegli di personalizzare il tuo bundle di applicazioni con l'opzione Personalizzato. Includi almeno Flink, Hadoop e Zeppelin nel tuo bundle.


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


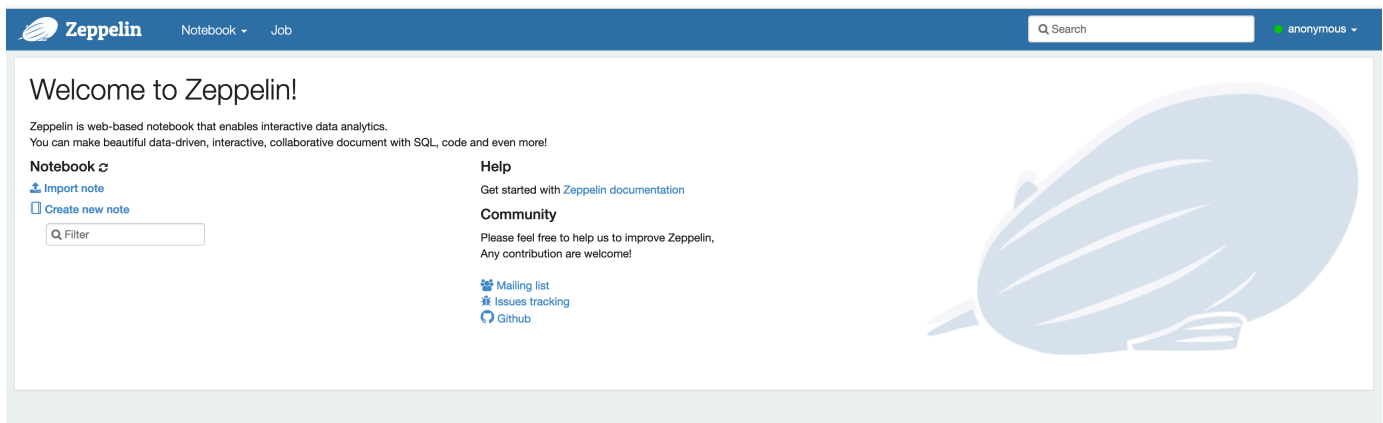
Custom


▼ 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. Crea il resto del cluster con le impostazioni che preferisci.
3. Una volta che il cluster è in esecuzione, selezionalo nella console per visualizzarne i dettagli e apri la scheda Applicazioni. Seleziona Zeppelin dalla sezione Interfacce utente dell'applicazione per aprire l'interfaccia Web di Zeppelin. Assicurati di aver configurato l'accesso all'interfaccia Web di Zeppelin con un tunnel SSH verso il nodo primario e una connessione proxy, come descritto in [Prerequisiti](#).



Zeppelin


Notebook - Job

Search

anonymous

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook 

[Import note](#)

[Create new note](#)

Filter

Help

Get started with [Zeppelin documentation](#)

Community

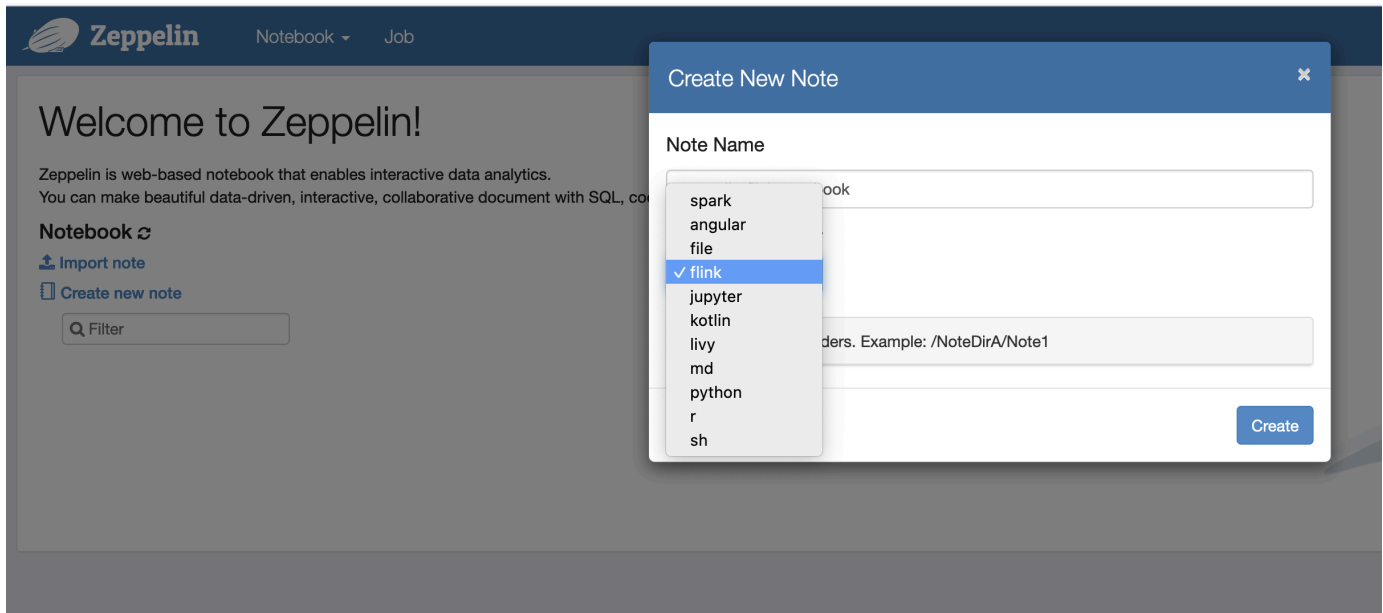
Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

[Mailing list](#)

[Issues tracking](#)

[Github](#)

4. Ora puoi creare una nuova nota in un notebook Zeppelin con Flink come interprete predefinito.



5. Fai riferimento ai seguenti esempi di codice che dimostrano come eseguire i processi Flink da un notebook Zeppelin.

Esecuzione dei processi Flink con Zeppelin-Flink su un cluster EMR

- Esempio 1, Scala in Flink

a) Esempio di WordCount in batch (SCALA)

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()
```

b) Esempio di WordCount in streaming (SCALA)

```
%flink

val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
```

```
.map(w => (w, 1))
.keyBy(0)
.sum(1)
.print

senv.execute()
```

Batch WordCount FLINK JOB FINISHED

```
%flink
val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
  .map(w => (w, 1))
  .groupByKey()
  .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)
```

Streaming WordCount FLINK JOB FINISHED

```
%flink
val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
  .map(w => (w, 1))
  .keyBy(0)
  .sum(1)
  .print()

senv.execute()

data: org.apache.flink.streaming.api.scala.DataStream[String] = org.apache.flink.streaming.api.scala.DataStream@202340f2
atastream.DataStreamSink@70be7b2f
warning: there was one deprecation warning; for details, enable ':setting -deprecation' or ':replay -deprecation'
res2: org.apache.flink.streaming.api.datastream.DataStreamSink[(String, Int)] = org.apache.flink.streaming.api.d
atastream.DataStreamSink@70be7b2f
res3: org.apache.flink.api.common.JobExecutionResult =
Program execution finished
Job with JobID a78d1a9b33c0fb71b8d53f00cda51030 has finished.
Job Runtime: 8485 ms

Took 12 sec. Last updated by anonymous at May 04 2023, 2:21:05 PM. (outdated)
```

- Esempio 2, Streaming SQL in Flink

```
%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>/glue-catalog-test/dbsamrat/t1/',
  'format' = 'csv'
);

INSERT INTO dummy_table SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2,
'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM dummy_table;
```

```
%flink.sql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table t1 (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://dbsamrat-emr-dev/glue-catalog-test/dbsamrat/t1/',
  'format' = 'csv'
);

INSERT INTO t1 SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2, 'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM t1;
```

FLINK JOB FINISHED

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

Took 20 sec. Last updated by anonymous at May 04 2023, 3:07:03 PM. (outdated)

• Esempio 3, Pyflink

```
%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. Scegli FLINK JOB nell'interfaccia utente di Zeppelin per accedere e visualizzare l'interfaccia utente Web di Flink.



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. Scegliendo FLINK JOB si accede alla console Web di Flink in un'altra scheda del browser.

The screenshot displays 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 shows:

- 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 currently shows "No Data".
- Completed Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, Status. It lists two jobs:

Job Name	Start Time	Duration	End Time	Tasks	Status
Flink Streaming Job	2023-05-04 14:20:56	8s	2023-05-04 14:21:04	2 / 2	FINISHED
Flink Java Job at Thu May 04 08:49:10 UTC 2023	2023-05-04 14:19:11	12s	2023-05-04 14:19:23	3 / 3	FINISHED

Cronologia dei rilasci di Flink

La tabella seguente indica la versione di Flink inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x o Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Flink

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-6.14.0	1,17.1-amzn-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,

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
		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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.12.0	1.17.0	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, delta-standalone-connectors

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-6.11.1	1.16.0	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, delta-standalone-connectors
emr-6.11.0	1.16.0	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, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-6.10.0	1.16.0	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-6.9.1	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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-5.6.0	1.2.1	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
emr-5.5.4	1.2.0	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, 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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
emr-5.5.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
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-resource manager, 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-resource manager, flink-client

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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-resource manager, flink-client
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
emr-5.2.2	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
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

Etichetta di rilascio di Amazon EMR	Versione di Flink	Componenti installati con Flink
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
emr-5.1.0	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

Ganglia

Il progetto open source Ganglia è un sistema scalabile e distribuito, progettato per monitorare i cluster e le griglie riducendo al minimo l'impatto sulle loro prestazioni. Quando abiliti Ganglia sul cluster, puoi generare report e visualizzare le prestazioni del cluster nel suo insieme, nonché esaminare le prestazioni di singole istanze nodi. Ganglia è anche configurato per acquisire e visualizzare i parametri Hadoop e Spark. Per ulteriori informazioni sul progetto open source Ganglia, visita la pagina all'indirizzo <http://ganglia.info/>.

Aperto l'interfaccia utente Web di Ganglia in un browser, puoi visualizzare una panoramica delle prestazioni del cluster, con grafici contenenti dettagli su carico, utilizzo della memoria, utilizzo della CPU e traffico di rete del cluster. Sotto alle statistiche del cluster, sono inclusi grafici per ogni singolo server del cluster.

La seguente tabella indica la versione di Ganglia inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Ganglia.

Per la versione dei componenti installati con Ganglia in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Ganglia per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-6.14.0	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

La seguente tabella indica la versione di Ganglia inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Ganglia.

Per la versione dei componenti installati con Ganglia in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Ganglia per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.36.1	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Argomenti

- [Creazione di un cluster con Ganglia](#)
- [Visualizzazione dei parametri di Ganglia](#)
- [Parametri Hadoop e Spark in Ganglia](#)
- [Cronologia delle versioni di Ganglia](#)

Creazione di un cluster con Ganglia

Per creare un cluster con Ganglia usando la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Scegli Crea cluster.
3. In Software configuration (Configurazione software), scegli All Applications (Tutte le applicazioni), Core Hadoop (Hadoop core) o Spark.

4. Procedere con la creazione del cluster con le configurazioni appropriate.

Per aggiungere Ganglia a un cluster tramite l'AWS CLI

Nell'AWS CLI, è possibile aggiungere Ganglia a un cluster utilizzando `create-cluster` con il parametro `--applications`. Se utilizzando il `--applications` si specifica solo Ganglia, questa sarà l'unica applicazione installata.

- Digitare il comando seguente per aggiungere Ganglia al momento della creazione di un cluster e sostituire *myKey* con il nome della coppia di chiavi di EC2.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name "Spark cluster with Ganglia" --release-label emr-5.36.1 \  
--applications Name=Spark Name=Ganglia \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3 --use-default-roles
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo master e le istanze rimanenti vengono avviate come nodi principali. Tutti i nodi utilizzano il tipo di istanza specificato nel comando.

Note

Se in precedenza non sono stati creati il ruolo del servizio EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

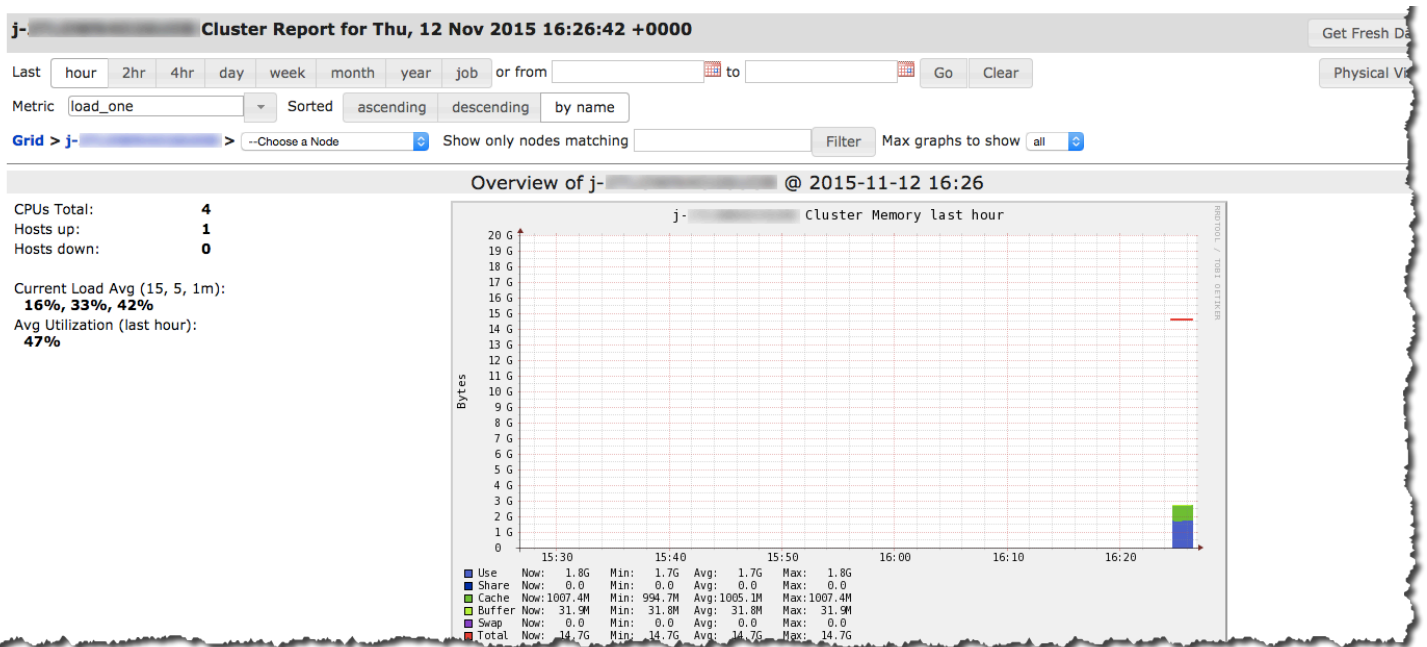
Per ulteriori informazioni sull'utilizzo dei comandi Amazon EMR in AWS CLI, consulta <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Visualizzazione dei parametri di Ganglia

Ganglia offre un'interfaccia utente basata sul Web che consente di visualizzare i parametri che l'applicazione raccoglie. Quando si esegue Ganglia su Amazon EMR, l'interfaccia Web viene eseguita sul nodo master e può essere visualizzata utilizzando l'inoltro porta, noto anche come creazione di un tunnel SSH. Per ulteriori informazioni sulla visualizzazione delle interfacce Web su Amazon EMR, consulta [Visualizzazione delle interfacce Web ospitate sui cluster EMR](#) nella Guida alla gestione di Amazon EMR.

Per visualizzare l'interfaccia Web Ganglia

1. Utilizzare SSH per eseguire il tunneling nel nodo master e creare una connessione sicura. Per informazioni su come creare un tunnel SSH per il nodo master, consulta la sezione relativa all'[Opzione 2, Parte 1: Impostazione di un tunnel SSH al nodo master tramite l'inoltro porta dinamico](#) nella Guida alla gestione di Amazon EMR.
2. Installa un browser Web con uno strumento proxy, ad esempio il plug-in FoxyProxy per Firefox, per creare un proxy SOCKS per domini del tipo `*ec2*.amazonaws.com*`. Per ulteriori informazioni, consulta la sezione relativa all'[Opzione 2, Parte 2: Configurazione delle impostazioni proxy per visualizzare siti Web ospitati sul nodo master](#) nella Guida alla gestione di Amazon EMR.
3. Con il proxy impostato e la connessione SSH aperta, è possibile visualizzare l'interfaccia utente di Ganglia aprendo una finestra del browser con `http://master-public-dns-name/ganglia/`, dove *master-public-dns-name* è l'indirizzo DNS pubblico del server master nel cluster EMR.



Parametri Hadoop e Spark in Ganglia

Ganglia riporta i parametri Hadoop per ogni istanza. Ai vari tipi di parametri sono aggiunti prefissi in base alla categoria: file system distribuito (dfs.*), macchina virtuale Java (jvm.*), MapReduce (mapred.*) e chiamate di procedura remota (rpc.*).

I parametri di Ganglia basati su YARN, quali Spark e Hadoop, non sono disponibili per EMR versione 4.4.0 e 4.5.0. Scegliere una versione successiva per utilizzare questi parametri.

I parametri di Ganglia per Spark generalmente hanno prefissi per l'ID applicazione YARN e DAGScheduler di Spark. I prefissi seguono quindi questa forma:

- DAGScheduler.*
- application_XXXXXXXXXX_XXXX.driver.*
- application_XXXXXXXXXX_XXXX.executor.*

Cronologia delle versioni di Ganglia

La seguente tabella indica la versione di Ganglia inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione Versione componente per il rilascio in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Ganglia

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-6.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.36.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-6.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-6.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.34.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.32.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.30.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.30.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.27.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.21.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.21.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.21.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.20.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.20.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.19.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.19.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.14.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.14.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
emr-5.13.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.12.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, hadoop-yarn-timeline-server, webserver
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-5.5.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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-resource-manager, 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-resource-manager, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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-resource-manager, webserver
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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-resource-manager, webserver
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
emr-4.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-4.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.0	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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
emr-4.7.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, webserver

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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
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

Etichetta di rilascio di Amazon EMR	Versione di Ganglia	Componenti installati con Ganglia
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](#) è un framework open source software Java che supporta l'elaborazione di enormi volumi di dati su un cluster di istanze. Può essere eseguito su una singola istanza o migliaia di istanze. Hadoop utilizza diversi modelli di elaborazione, ad esempio MapReduce e Tez, per distribuire l'elaborazione su più istanze; utilizza inoltre un file system distribuito denominato HDFS per archiviare i dati su più istanze. Hadoop monitora lo stato delle istanze del cluster ed è in grado di recuperarle in caso di errore di uno o più nodi. In questo modo, Hadoop fornisce una maggiore elaborazione e capacità di storage, nonché l'elevata disponibilità. Per ulteriori informazioni, consulta la [documentazione Hadoop](#).

La seguente tabella indica la versione di Hadoop inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Hadoop.

Per la versione dei componenti installati con Hadoop in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hadoop per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.14.0	Hadoop 3.3.3	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-resource-manager, hadoop-yarn-timeline-server

La seguente tabella indica la versione di Hadoop inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Hadoop.

Per la versione dei componenti installati con Hadoop in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hadoop per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.36.1	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

A partire da Amazon EMR rilascio 5.18.0, è possibile utilizzare i repository degli artefatti di Amazon EMR per creare il codice di processo secondo le versioni esatte di librerie e dipendenze disponibili con rilasci Amazon EMR specifici. Per ulteriori informazioni, consulta [Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR](#).

Argomenti

- [Configura Hadoop](#)
- [Crittografia trasparente in HDFS su Amazon EMR](#)
- [Creazione ed esecuzione di un'applicazione Hadoop](#)
- [Attiva il riconoscimento non uniforme degli accessi alla memoria per i container YARN](#)
- [Cronologia delle versioni di Hadoop](#)

Configura Hadoop

Le seguenti sezioni forniscono le impostazioni di configurazione predefinite per i daemon Hadoop, le attività e gli HDFS.

Argomenti

- [Configurazione attività](#)
- [Impostazioni di configurazione Daemon Hadoop](#)
- [Configurazione HDFS](#)

Configurazione attività

È possibile impostare variabili di configurazione per ottimizzare le prestazioni dei lavori di MapReduce. Questa sezione fornisce i valori predefiniti per le impostazioni importanti. I valori predefiniti variano in base al tipo di istanza EC2 del nodo utilizzato nel cluster. HBase è disponibile quando si utilizza Amazon EMR versione 4.6.0 e versioni successive. Quando si installa HBase vengono utilizzati valori predefiniti diversi. Questi valori sono forniti insieme alle impostazioni predefinite iniziali.

Hadoop 2 usa due parametri, `mapreduce.map.java.opts` e `mapreduce.reduce.java.opts`, per configurare la memoria per mappare e ridurre JVM rispettivamente. Queste sostituiscono l'opzione di configurazione singola `mapreduce.map.java.opts` delle versioni Hadoop precedenti.

Analogamente, `mapred.job.jvm.num.tasks` sostituisce `mapred.job.reuse.jvm.num.tasks` in Hadoop 2.7.2 e versioni successive. Amazon EMR imposta questo valore a 20 indipendentemente dal tipo di istanza di EC2. È possibile ignorare questa impostazione utilizzando la classificazione di `mapred-site` configurazione. Impostando un valore di `-1` si indica che una JVM può essere utilizzata per un numero infinito di attività all'interno di un singolo processo, e il valore `1` indica che una nuova JVM viene generata dinamicamente per ciascuna operazione.

Ad esempio, per impostare il valore di `mapred.job.jvm.num.tasks` su `-1`, è possibile creare un file con il seguente contenuto:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

Quando si utilizza il comando `create-cluster` o il comando `modify-instance-groups` da AWS CLI, è possibile fare riferimento al file di configurazione JSON. In questo esempio, il file di configurazione viene salvato come `myConfig.json` e memorizzato in Amazon S3.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge \  
--instance-count 3 --applications Name=Hadoop --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json \  
--use-default-roles
```

È possibile modificare i valori predefiniti elencati di seguito utilizzando la classificazione di configurazione `mapred-site` nello stesso modo e impostare più valori e classificazioni di configurazione utilizzando un unico file JSON. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Valori predefiniti per le impostazioni di configurazione delle attività

Tipi di istanza

- [Istanze c1](#)
- [Istanze c3](#)
- [Istanze c4](#)
- [Istanze c5](#)
- [Istanze c5a](#)
- [Istanze c5ad](#)

- [Istanze c5d](#)
- [Istanze c5n](#)
- [Istanze c6a](#)
- [Istanze c6g](#)
- [Istanze c6gd](#)
- [Istanze c6gn](#)
- [Istanze c6i](#)
- [Istanze c6id](#)
- [Istanze c6in](#)
- [Istanze c7g](#)
- [Istanze c7gd](#)
- [Istanze c7gn](#)
- [Istanze c7i](#)
- [Istanze d2](#)
- [Istanze d3](#)
- [Istanze d3en](#)
- [Istanze g3](#)
- [Istanze g3s](#)
- [Istanze g4dn](#)
- [Istanze g5](#)
- [Istanze h1](#)
- [Istanze i2](#)
- [Istanze i3](#)
- [Istanze i3en](#)
- [istanze i4g](#)
- [Istanze i4i](#)
- [Istanze im4gn](#)
- [Istanze is4gen](#)

- [Istanze m1](#)
- [Istanze m2](#)
- [Istanze m3](#)
- [Istanze m4](#)
- [Istanze m5](#)
- [Istanze m5a](#)
- [Istanze m5ad](#)
- [Istanze m5d](#)
- [Istanze m5dn](#)
- [Istanze m5n](#)
- [Istanze m5zn](#)
- [Istanze m6a](#)
- [Istanze m6g](#)
- [Istanze m6gd](#)
- [Istanze m6i](#)
- [Istanze m6id](#)
- [Istanze m6idn](#)
- [Istanze m6in](#)
- [Istanze m7a](#)
- [Istanze m7g](#)
- [Istanze m7gd](#)
- [Istanze m7i](#)
- [Istanze m7i-flex](#)
- [Istanze p2](#)
- [Istanze p3](#)
- [Istanze p5](#)
- [Istanze r3](#)
- [Istanze r4](#)
- [Istanze r5](#)

- [Istanze r5a](#)
- [Istanze r5ad](#)
- [Istanze r5b](#)
- [Istanze r5d](#)
- [Istanze r5dn](#)
- [Istanze r5n](#)
- [Istanze r6a](#)
- [Istanze r6g](#)
- [Istanze r6gd](#)
- [Istanze r6i](#)
- [Istanze r6id](#)
- [Istanze r6idn](#)
- [Istanze r6in](#)
- [Istanze r7a](#)
- [Istanza r7g](#)
- [Istanze r7gd](#)
- [Istanze r7iz](#)
- [Istanze x1](#)
- [Istanze x1e](#)
- [Istanze x2gd](#)
- [istanze x2idn](#)
- [Istanze x2iedn](#)
- [Istanze z1d](#)

Istanze c1

c1.medium

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1.024	512

c1.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx864m</code>	<code>-Xmx864m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	1.024	1.024
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	2560

Opzione di configurazione	Valore predefinito	Con HBase installato
yarn.nodemanager.resource.memory-mb	5120	2560

Istanze c3

c3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
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

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

c3.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1331m</code>	<code>-Xmx1331m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2662m</code>	<code>-Xmx2662m</code>
<code>mapreduce.map.memory.mb</code>	1664	1664
<code>mapreduce.reduce.memory.mb</code>	3328	3328
<code>yarn.app.mapreduce.am.resource.mb</code>	3328	3328
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	53248	26624
<code>yarn.nodemanager.resource.memory-mb</code>	53248	26624

Istanze c4

c4.large

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx717m</code>	<code>-Xmx717m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1434m</code>	<code>-Xmx1434m</code>

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c4.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

c4.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c5

c5.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

c5.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

c5.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

c5.9xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

c5.18xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource.memory-mb</code>	139264	30960

c5.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze c5a

c5a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c5a.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

c5a.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c5ad

c5ad.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c5ad.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c5ad.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c5ad.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c5ad.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c5d

c5d.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

c5d.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

c5d.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

c5d.9xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze c5n

c5n.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c6a

c6a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

c6a.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6a.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c6a.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

c6a.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1527m</code>	<code>-Xmx1527m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3054m</code>	<code>-Xmx3054m</code>
<code>mapreduce.map.memory.mb</code>	1909	1909
<code>mapreduce.reduce.memory.mb</code>	3818	3818
<code>yarn.app.mapreduce.am.resource.mb</code>	3818	3818
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

Istanze c6g

c6g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6g.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6g.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c6gd

c6gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c6gn

c6gn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c6i

c6i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c6id

c6id.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6id.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6id.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze c6in

c6in.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6in.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6in.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6in.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6in.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c6in.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Istanze c7g

c7g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7g.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7g.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c7gd

c7gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7gd.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c7gn

c7gn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7gn.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7gn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7gn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze c7i

c7i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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>	1408	1408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

c7i.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

Istanze d2

d2.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

d2.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

d2.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

d2.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2417m</code>	<code>-Xmx2417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4834m</code>	<code>-Xmx4834m</code>
<code>mapreduce.map.memory.mb</code>	3021	3021
<code>mapreduce.reduce.memory.mb</code>	6042	6042
<code>yarn.app.mapreduce.am.resource.mb</code>	6042	6042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30194
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30194

Istanze d3

d3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

d3.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

d3.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

d3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze d3en

d3en.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

d3en.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

d3en.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

d3en.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

d3en.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Istanze g3

g3.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

g3.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

g3.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze g3s

g3s.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze g4dn

g4dn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

g4dn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

g4dn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze g5

g5.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

g5.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

g5.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

g5.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

g5.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze h1

h1.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

h1.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

h1.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze i2

i2.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i2.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i2.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Istanze i3

i3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

i3.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i3.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i3.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze i3en

i3en.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

i3en.6xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6281m</code>	<code>-Xmx6281m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12562m</code>	<code>-Xmx12562m</code>
<code>mapreduce.map.memory.mb</code>	7851	7851
<code>mapreduce.reduce.memory.mb</code>	15702	15702
<code>yarn.app.mapreduce.am.resource.mb</code>	15702	15702
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

i3en.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

istanze i4g

i4g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i4g.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i4i.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

i4i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze i4i

i4i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i4i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i4i.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

i4i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

i4i.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze im4gn

im4gn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

im4gn.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

im4gn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Istanze is4gen

is4gen.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3514m</code>	<code>-Xmx3514m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx7028m</code>	<code>-Xmx7028m</code>
<code>mapreduce.map.memory.mb</code>	4393	4393
<code>mapreduce.reduce.memory.mb</code>	8786	8786
<code>yarn.app.mapreduce.am.resource.mb</code>	8786	8786
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	17572.12	8786.06
<code>yarn.nodemanager.resource.memory-mb</code>	17572.12	8786.06

is4gen.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3866m</code>	<code>-Xmx3866m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx7732m</code>	<code>-Xmx7732m</code>
<code>mapreduce.map.memory.mb</code>	4832	4832
<code>mapreduce.reduce.memory.mb</code>	9664	9664
<code>yarn.app.mapreduce.am.resource.mb</code>	9664	9664

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	38656	19328
<code>yarn.nodemanager.resource.memory-mb</code>	38656	19328

is4gen.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4275m</code>	<code>-Xmx4275m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx8550m</code>	<code>-Xmx8550m</code>
<code>mapreduce.map.memory.mb</code>	5344	5344
<code>mapreduce.reduce.memory.mb</code>	10688	10688
<code>yarn.app.mapreduce.am.resource.mb</code>	10688	10688
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32064
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32064

is4gen.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m1

m1.small

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource.memory-mb</code>	1.024	512

m1.medium

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx768m</code>	<code>-Xmx768m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	1.024	1.024
<code>yarn.app.mapreduce.am.resource.mb</code>	1.024	1.024
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	1.024
<code>yarn.nodemanager.resource.memory-mb</code>	2048	1.024

m1.large

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx1024m	-Xmx1024m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	1536	1536
yarn.app.mapreduce.am.resource.mb	1536	1536
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	3072	2560
yarn.nodemanager.resource.memory-mb	5120	2560

m1.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

Istanze m2

m2.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx864m</code>	<code>-Xmx864m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	1.024	1.024
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	7168	7168
<code>yarn.nodemanager.resource.memory-mb</code>	14336	7168

m2.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	30720
<code>yarn.nodemanager.resource.memory-mb</code>	61440	30720

Istanze m3

m3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

m3.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m4

m4.large

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

m4.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m4.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

m4.10xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx1557m</code>	<code>-Xmx1557m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3114m</code>	<code>-Xmx3114m</code>
<code>mapreduce.map.memory.mb</code>	1946	1946
<code>mapreduce.reduce.memory.mb</code>	3892	3892
<code>yarn.app.mapreduce.am.resource.mb</code>	3892	3892
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	155648	31104
<code>yarn.nodemanager.resource.memory-mb</code>	155648	31104

m4.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m5

m5.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m5.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

m5.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

m5.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m5a

m5a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m5a.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

m5a.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5a.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

m5a.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3209m</code>	<code>-Xmx3209m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6418m</code>	<code>-Xmx6418m</code>
<code>mapreduce.map.memory.mb</code>	4011	4011
<code>mapreduce.reduce.memory.mb</code>	8022	8022
<code>yarn.app.mapreduce.am.resource.mb</code>	8022	8022
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

Istanze m5ad

m5ad.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

m5ad.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5ad.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m5ad.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972 m.	-Xmx5972 m.
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m5ad.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5ad.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m5d

m5d.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource-memory-mb</code>	12288	6144

m5d.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource-memory-mb</code>	24576	12288

m5d.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5d.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

m5d.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m5dn

m5dn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m5dn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m5dn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m5dn.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5dn.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m5n

m5n.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

m5n.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource-memory-mb</code>	54272	27136

m5n.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972 m.	-Xmx5972 m.
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m5n.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5n.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m5zn

m5zn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

m5zn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5zn.3xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

m5zn.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972 m.</code>	<code>-Xmx5972 m.</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

Istanze m6a

m6a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6a.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972 m.	-Xmx5972 m.
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m6a.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6a.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

m6a.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m6g

m6g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6g.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6g.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6g.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Istanze m6gd

m6gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

m6gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6gd.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6gd.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6gd.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Istanze m6i

m6i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6i.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6i.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6i.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m6id

m6id.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6id.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6id.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m6id.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6id.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m6id.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m6idn

m6idn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6idn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6idn.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m6idn.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972 m.</code>	<code>-Xmx5972 m.</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m6idn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m6idn.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m6in

m6in.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6in.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

m6in.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6in.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6in.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m7a

m7a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7a.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m7g

m7g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7g.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7g.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7g.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m7gd

m7gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7gd.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7gd.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-XMx5972 m.	-XMx5972 m.
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze m7i

m7i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7i.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7i.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7i.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m7i.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m7i.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze m7i-flex

m7i-flex.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7i-flex.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m7i-flex.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze p2

p2.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx21708m</code>	<code>-Xmx21708m</code>
<code>mapreduce.map.memory.mb</code>	13568	13568
<code>mapreduce.reduce.memory.mb</code>	27136	27136
<code>yarn.app.mapreduce.am.resource.mb</code>	27136	27136
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

p2.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

p2.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	23168
<code>yarn.nodemanager.resource.memory-mb</code>	741376	23168

Istanze p3

p3.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

p3.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

p3.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze p5

p5.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx8294m</code>	<code>-Xmx8294m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx16588m</code>	<code>-Xmx16588m</code>
<code>mapreduce.map.memory.mb</code>	10368	10368
<code>mapreduce.reduce.memory.mb</code>	20736	20736
<code>yarn.app.mapreduce.am.resource.mb</code>	20736	20736
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	20736
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	20736

Istanze r3

r3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r3.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r3.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r4

r4.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r4.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r4.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r4.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Istanze r5

r5.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

r5.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

r5.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6451m</code>	<code>-Xmx6451m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12902m</code>	<code>-Xmx12902m</code>
<code>mapreduce.map.memory.mb</code>	8064	8064
<code>mapreduce.reduce.memory.mb</code>	16128	16128
<code>yarn.app.mapreduce.am.resource.mb</code>	16128	16128
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	516096	32256
<code>yarn.nodemanager.resource.memory-mb</code>	516096	32256

r5.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r5a

r5a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

r5a.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

r5a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

r5a.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5a.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r5ad

r5ad.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5ad.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5ad.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r5ad.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	499712	31232
<code>yarn.nodemanager.resource.memory-mb</code>	499712	31232

r5ad.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r5b

r5b.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r5b.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5b.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5b.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r5b.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5b.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r5d

r5d.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

r5d.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

r5d.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

r5d.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5d.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r5dn

r5dn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5dn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5dn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r5dn.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5dn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5dn.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r5n

r5n.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r5n.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5n.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r5n.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5n.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r6a

r6a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6a.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

r6a.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6a.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6a.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6a.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

r6a.48xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r6g

r6g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource-memory-mb</code>	23424	11712

r6g.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource-memory-mb</code>	54272	27136

r6g.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6g.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6g.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r6gd

r6gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6gd.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6gd.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6gd.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r6i

r6i.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6i.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6i.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6i.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6i.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6i.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6029m</code>	<code>-Xmx6029m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.map.memory.mb</code>	7536	7536
<code>mapreduce.reduce.memory.mb</code>	15072	15072
<code>yarn.app.mapreduce.am.resource.mb</code>	15072	15072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

Istanze r6id

r6id.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6id.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6id.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6id.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6id.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r6id.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6id.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r6idn

r6idn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6idn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6idn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6idn.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r6idn.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6idn.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Istanze r6in

r6in.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6in.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6in.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6in.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6in.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6in.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanze r7a

r7a.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7a.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Istanza r7g

r7g.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7g.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7g.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7g.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Istanze r7gd

r7gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7gd.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7gd.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7gd.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Istanze r7iz

r7iz.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7iz.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7iz.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

r7iz.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

r7iz.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7iz.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r7iz.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

Istanze x1

x1.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
mapreduce.map.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.java.opts	-Xmx24116m	-Xmx24116m
mapreduce.map.memory.mb	15072	15072
mapreduce.reduce.memory.mb	30144	30144
yarn.app.mapreduce.am.resource.mb	30144	30144

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx12109m</code>	<code>-Xmx12109m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24218m</code>	<code>-Xmx24218m</code>
<code>mapreduce.map.memory.mb</code>	15136	15136
<code>mapreduce.reduce.memory.mb</code>	30272	30272
<code>yarn.app.mapreduce.am.resource.mb</code>	30272	30272
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1937408	30272
<code>yarn.nodemanager.resource.memory-mb</code>	1937408	30272

Istanze x1e

x1e.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx22682m</code>	<code>-Xmx22682m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx45364m</code>	<code>-Xmx45364m</code>
<code>mapreduce.map.memory.mb</code>	28352	28352
<code>mapreduce.reduce.memory.mb</code>	56704	56704
<code>yarn.app.mapreduce.am.resource.mb</code>	56704	56704
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	113408	0
<code>yarn.nodemanager.resource.memory-mb</code>	113408	0

x1e.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx23501m</code>	<code>-Xmx23501m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx47002m</code>	<code>-Xmx47002m</code>
<code>mapreduce.map.memory.mb</code>	29376	29376
<code>mapreduce.reduce.memory.mb</code>	58752	58752
<code>yarn.app.mapreduce.am.resource.mb</code>	58752	58752

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	235008	0
<code>yarn.nodemanager.resource.memory-mb</code>	235008	0

x1e.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx23910m</code>	<code>-Xmx23910m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx47820m</code>	<code>-Xmx47820m</code>
<code>mapreduce.map.memory.mb</code>	29888	29888
<code>mapreduce.reduce.memory.mb</code>	59776	59776
<code>yarn.app.mapreduce.am.resource.mb</code>	59776	59776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	478208	0
<code>yarn.nodemanager.resource.memory-mb</code>	478208	0

x1e.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	0
yarn.nodemanager.resource.memory-mb	964608	0

x1e.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1937408	0
<code>yarn.nodemanager.resource.memory-mb</code>	1937408	0

x1e.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx24269m</code>	<code>-Xmx24269m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx48538m</code>	<code>-Xmx48538m</code>
<code>mapreduce.map.memory.mb</code>	30336	30336
<code>mapreduce.reduce.memory.mb</code>	60672	60672
<code>yarn.app.mapreduce.am.resource.mb</code>	60672	60672
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	3883008	0
<code>yarn.nodemanager.resource.memory-mb</code>	3883008	0

Istanze x2gd

x2gd.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx21708m</code>	<code>-Xmx21708m</code>
<code>mapreduce.map.memory.mb</code>	13568	13568
<code>mapreduce.reduce.memory.mb</code>	27136	27136
<code>yarn.app.mapreduce.am.resource.mb</code>	27136	27136
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

x2gd.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx23348m</code>	<code>-Xmx23348m</code>
<code>mapreduce.map.memory.mb</code>	14592	14592
<code>mapreduce.reduce.memory.mb</code>	29184	29184
<code>yarn.app.mapreduce.am.resource.mb</code>	29184	29184

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

x2gd.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx12083m</code>	<code>-Xmx12083m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24166m</code>	<code>-Xmx24166m</code>
<code>mapreduce.map.memory.mb</code>	15104	15104
<code>mapreduce.reduce.memory.mb</code>	30208	30208
<code>yarn.app.mapreduce.am.resource.mb</code>	30208	30208
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

x2gd.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

x2gd.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30906
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30906

x2gd.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24780m</code>	<code>-Xmx24780m</code>
<code>mapreduce.map.memory.mb</code>	15488	15488
<code>mapreduce.reduce.memory.mb</code>	30976	30976
<code>yarn.app.mapreduce.am.resource.mb</code>	30976	30976
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

istanze x2idn

x2idn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31030
<code>yarn.nodemanager.resource.memory-mb</code>	1490944	31030

x2idn.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx12442m</code>	<code>-Xmx12442m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24884m</code>	<code>-Xmx24884m</code>
<code>mapreduce.map.memory.mb</code>	15552	15552
<code>mapreduce.reduce.memory.mb</code>	31104	31104
<code>yarn.app.mapreduce.am.resource.mb</code>	31104	31104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	31104
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	31104

Istanze x2iedn

x2iedn.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	0
yarn.nodemanager.resource.memory-mb	116736	0

x2iedn.2xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	0
<code>yarn.nodemanager.resource.memory-mb</code>	241664	0

x2iedn.4xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49152m</code>	<code>-Xmx49152m</code>
<code>mapreduce.map.memory.mb</code>	30720	30720
<code>mapreduce.reduce.memory.mb</code>	61440	61440
<code>yarn.app.mapreduce.am.resource.mb</code>	61440	61440
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	0
<code>yarn.nodemanager.resource.memory-mb</code>	491520	0

x2iedn.8xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	0
yarn.nodemanager.resource.memory-mb	991232	0

x2iedn.16xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	0
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	0

x2iedn.24xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx24918m</code>	<code>-Xmx24918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49836m</code>	<code>-Xmx49836m</code>
<code>mapreduce.map.memory.mb</code>	31147	31147
<code>mapreduce.reduce.memory.mb</code>	62294	62294
<code>yarn.app.mapreduce.am.resource.mb</code>	62294	62294
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	2990080	-32
<code>yarn.nodemanager.resource.memory-mb</code>	2990080	-32

x2iedn.32xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3989504	0
yarn.nodemanager.resource.memory-mb	3989504	0

Istanze z1d

z1d.xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

z1d.3xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

z1d.6xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
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

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

z1d.12xlarge

Opzione di configurazione	Valore predefinito	Con HBase installato
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

Impostazioni di configurazione Daemon Hadoop

Le impostazioni di daemon Hadoop sono diverse, in funzione del tipo di istanza EC2 che si avvale di un nodo del cluster. Le seguenti tabelle elencano le impostazioni di configurazione predefinite per ogni tipo di istanza EC2.

Per personalizzare queste impostazioni, utilizza la classificazione di configurazione `hadoop-env`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Tipi di istanza

- [Istanze c1](#)
- [Istanze c3](#)
- [Istanze c4](#)
- [Istanze c5](#)
- [Istanze c5a](#)
- [Istanze c5ad](#)
- [Istanze c5d](#)
- [Istanze c5n](#)
- [Istanze c6a](#)
- [Istanze c6g](#)
- [Istanze c6gd](#)
- [Istanze c6gn](#)
- [Istanze c6i](#)
- [Istanze c6id](#)
- [Istanze c6in](#)
- [Istanze c7g](#)
- [Istanze c7gd](#)
- [Istanze c7gn](#)
- [Istanze d2](#)
- [Istanze d3](#)
- [Istanze d3en](#)
- [Istanze g2](#)

- [Istanze g3](#)
- [Istanze g3s](#)
- [Istanze g4dn](#)
- [Istanze g5](#)
- [Istanze h1](#)
- [Istanze i2](#)
- [Istanze i3](#)
- [Istanze i3en](#)
- [istanze i4g](#)
- [Istanze i4i](#)
- [Istanze im4gn](#)
- [Istanze is4gen](#)
- [Istanze m1](#)
- [Istanze m2](#)
- [Istanze m3](#)
- [Istanze m4](#)
- [Istanze m5](#)
- [Istanze m5a](#)
- [Istanze m5ad](#)
- [Istanze m5d](#)
- [Istanze m5dn](#)
- [Istanze m5n](#)
- [Istanze m5zn](#)
- [Istanze m6a](#)
- [Istanze m6g](#)
- [Istanze m6gd](#)
- [Istanze m6i](#)
- [Istanze m6id](#)
- [Istanze m6idn](#)

- [Istanze m6in](#)
- [Istanze m7g](#)
- [Istanze m7gd](#)
- [Istanze m7i](#)
- [Istanze m7i-flex](#)
- [Istanze p2](#)
- [Istanze p3](#)
- [Istanze p5](#)
- [Istanze r3](#)
- [Istanze r4](#)
- [Istanze r5](#)
- [Istanze r5a](#)
- [Istanze r5ad](#)
- [Istanze r5b](#)
- [Istanze r5d](#)
- [Istanze r5dn](#)
- [Istanze r5n](#)
- [Istanze r6a](#)
- [Istanze r6g](#)
- [Istanze r6gd](#)
- [Istanze r6i](#)
- [Istanze r6id](#)
- [Istanze r6idn](#)
- [Istanze r6in](#)
- [Istanza r7g](#)
- [Istanze r7gd](#)
- [Istanze x1](#)
- [Istanze x1e](#)
- [Istanze x2gd](#)

- [istanze x2idn](#)
- [Istanze x2iedn](#)
- [Istanze z1d](#)

Istanze c1

c1.medium

Parametro	Valore
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

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768

Parametro	Valore
HADOOP_DATANODE_HEAPSIZE	384

Istanze c3

c3.xlarge

Parametro	Valore
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

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740

Parametro	Valore
HADOOP_DATANODE_HEAPSIZE	757

c3.4xlarge

Parametro	Valore
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

Parametro	Valore
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

Istanze c4

c4.large

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c5

c5.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c5a

c5a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c5ad

c5ad.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c5d

c5d.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c5n

c5n.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6a

c6a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6g

c6g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6gd

c6gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6gn

c6gn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6i

c6i.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6id

c6id.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c6in

c6in.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c7g

c7g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c7gd

c7gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze c7gn

c7gn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze d2

d2.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze d3

d3.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze d3en

d3en.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze g2

g2.2xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	2304
HADOOP_DATANODE_HEAPSIZE	384

Istanze g3

g3.4xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

g3.8xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

g3.16xlarge

Parametro	Valore
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

Istanze g3s

g3s.xlarge

Parametro	Valore
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

Istanze g4dn

g4dn.xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

g4dn.2xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

g4dn.4xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

g4dn.8xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

g4dn.12xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

g4dn.16xlarge

Parametro	Valore
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

Istanze g5

g5.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze h1

h1.2xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze i2

i2.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze i3

i3.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze i3en

i3en.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

istanze i4g

i4g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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.16xlarge

Parametro	Valore
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

Istanze i4i

i4i.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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.16xlarge

Parametro	Valore
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.32xlarge

Parametro	Valore
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

Istanze im4gn

im4gn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze is4gen

is4gen.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m1

m1.small

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m2

m2.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m3

m3.xlarge

Parametro	Valore
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

Parametro	Valore
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

Istanze m4

m4.large

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5

m5.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5a

m5a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5ad

m5ad.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5d

m5d.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5dn

m5dn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5n

m5n.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m5zn

m5zn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6a

m6a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6g

m6g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6gd

m6gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6i

m6i.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6id

m6id.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6idn

m6idn.xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6idn.2xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m6in

m6in.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6in.8xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m7g

m7g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m7gd

m7gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m7i

m7i.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze m7i-flex

m7i-flex.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze p2

p2.xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p2.8xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

p2.16xlarge

Parametro	Valore
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

Istanze p3

p3.2xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p3.8xlarge

Parametro	Valore
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

p3.16xlarge

Parametro	Valore
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

Istanze p5

p5.48xlarge

Parametro	Valore
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

Istanze r3

r3.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r4

r4.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5

r5.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5a

r5a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5ad

r5ad.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5b

r5b.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5d

r5d.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5dn

r5dn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r5n

r5n.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6a

r6a.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6g

r6g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6gd

r6gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6i

r6i.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6id

r6id.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6idn

r6idn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r6in

r6in.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanza r7g

r7g.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze r7gd

r7gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze x1

x1.16xlarge

Parametro	Valore
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

Parametro	Valore
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

Istanze x1e

x1e.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze x2gd

x2gd.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

istanze x2idn

x2idn.16xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze x2iedn

x2iedn.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Istanze z1d

z1d.xlarge

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Parametro	Valore
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

Configurazione HDFS

La tabella seguente descrive i parametri predefiniti del file di sistema distribuito Hadoop (HDFS) e le relative impostazioni. Questi valori possono essere modificati utilizzando la classificazione di configurazione `hdfs-site`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Warning

1. L'impostazione di `dfs.replication` su 1 per i cluster con meno di quattro nodi può causare la perdita di dati HDFS in caso di disattivazione anche di un singolo nodo. Se il cluster dispone di spazio di archiviazione HDFS, per evitare la perdita di dati è preferibile configurarlo con almeno quattro nodi principali per i carichi di lavoro di produzione.
2. Amazon EMR non consente ai cluster di dimensionare i nodi principali al di sotto di `dfs.replication`. Ad esempio, se `dfs.replication = 2`, il numero minimo di nodi principali è 2.
3. Quando utilizzi il dimensionamento gestito, il dimensionamento automatico o scegli di dimensionare manualmente il cluster, ti consigliamo di impostare `dfs.replication` su 2 o su un valore superiore.

Parametro	Definizione	Valore predefinito
<code>dfs.block.size</code>	La dimensione dei blocchi HDFS. Quando si opera su dati memorizzati in HDFS, la dimensione di split è generalmente la dimensione di un blocco HDFS. Un numero maggiore garantisce una minore granularità delle attività, ma anche una minore sollecitazione del cluster NameNode.	134217728 (128 MB)
<code>dfs.replication</code>	Il numero di copie di ogni blocco da conservare per una maggiore durata. Amazon EMR imposta questo valore in base al numero di nodi principali con cui viene fornito il cluster. Modifica il valore in base alle tue esigenze. Per sovrascrivere il valore di default, è possibile usare la classificazione <code>hdfs-site</code> .	1 per cluster sottoposti a provisioning con meno di quattro nodi principali 2 per cluster sottoposti a

Parametro	Definizione	Valore predefinito
		provisioning con meno di dieci nodi principali
		3 per tutti gli altri cluster

Crittografia trasparente in HDFS su Amazon EMR

La crittografia trasparente è implementata attraverso l'utilizzo di zone di crittografia di HDFS, che sono i percorsi HDFS definiti dall'utente. Ogni zona di crittografia ha una propria chiave, che viene memorizzata nel server di chiavi specificato utilizzando la classificazione di configurazione `hdfs-site`.

A partire dalla versione Amazon EMR 4.8.0, puoi utilizzare le configurazioni di sicurezza di Amazon EMR per configurare più facilmente impostazioni di crittografia di dati per cluster. Le configurazioni di sicurezza offrono impostazioni che assicurano la sicurezza dei dati in transito e a riposo nei volumi di archiviazione di Amazon Elastic Block Store (Amazon EBS) e in EMRFS su Amazon S3. Per ulteriori informazioni, consulta [Crittografia dei dati in transito e a riposo](#) nella Guida alla gestione di Amazon EMR.

Amazon EMR usa il KMS Hadoop per impostazione predefinita; tuttavia, è possibile usare un altro KMS che implementa l'operazione API del KeyProvider. Ogni file in una zona di crittografia HDFS ha le proprie chiavi di crittografia dei dati univoche, che sono crittografate in base alla chiave della zona di crittografia. I dati HDFS vengono crittografati da punto a punto (memorizzati e in transito) quando vengono scritti in una zona di crittografia perché le attività di crittografia e decrittografia avvengono solo nel client.

Non è possibile spostare file tra zone di crittografia o da una zona di crittografia a percorsi non crittografati.

Il client NameNode e HDFS interagiscono con il KMS Hadoop (o un KMS alternativo configurato dall'utente) tramite l'operazione API KeyProvider. Il KMS è responsabile della memorizzazione delle chiavi di crittografia nel deposito chiavi di supporto. Inoltre, Amazon EMR include la policy di forza illimitata JCE, in modo da poter creare chiavi della lunghezza desiderata.

Per ulteriori informazioni, consulta [Crittografia trasparente in HDFS](#) nella documentazione di Hadoop.

Note

In Amazon EMR, KMS su HTTPS non è abilitato per impostazione predefinita con Hadoop KMS. Per ulteriori informazioni su come abilitare KMS su HTTPS, consulta la [documentazione di KMS Hadoop](#).

Configurazione della crittografia trasparente HDFS

Puoi configurare la crittografia trasparente in Amazon EMR mediante la creazione di chiavi e aggiungendo zone di crittografia. Ci sono diversi modi per farlo:

- Utilizzo dell'operazione API di configurazione di Amazon EMR quando si crea un cluster
- Usare una fase Hadoop JAR con `command-runner.jar`
- Accedendo al nodo master del cluster Hadoop e utilizzando i client della riga di comando `hadoop key` e `hdfs crypto`
- Utilizzando l'API REST per Hadoop KMS e HDFS

Per ulteriori informazioni sulle API REST, consulta la relativa documentazione per Hadoop KMS e HDFS.

Per creare zone di crittografia e le relative chiavi durante la creazione di cluster utilizzando la CLI

La classificazione `hdfs-encryption-zones` nell'operazione API di configurazione consente di specificare un nome di chiave e una zona di crittografia quando si crea un cluster. Amazon EMR crea questa chiave in Hadoop KMS sul cluster e configura l'area di crittografia.

- Creare un cluster con il seguente comando.

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

 Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).


myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

Per creare manualmente le zone di crittografia e le relative chiavi sul nodo master

1. Avviare il cluster utilizzando una versione di Amazon EMR superiore a 4.1.0.
2. Connettersi al nodo principale del cluster con SSH.
3. Creare una chiave all'interno di Hadoop KMS.

```
$ 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 richiede che i nomi delle chiavi chiave siano minuscoli. Se si utilizza una chiave con caratteri maiuscoli, il cluster fallirà durante l'avvio.

4. Creare il percorso della zona di crittografia in HDFS.

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Rendere il percorso HDFS una zona di crittografia utilizzando la chiave creata.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

Per creare manualmente le zone di crittografia e le relative usando la AWS CLI

- Aggiungere i passaggi per creare manualmente le chiavi KMS e le zone di crittografia con il seguente comando.

```
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

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Considerazioni per la crittografia trasparente HDFS

Una best practice è quella di creare una zona di crittografia per ogni applicazione in cui possono scrivere file. Inoltre, è possibile crittografare tutti gli HDFS utilizzando la classificazione delle zone di cifratura `hdfs-encryption` nell'API di configurazione e specificare il percorso di root (`/`) come zona di cifratura.

Server di gestione delle chiavi Hadoop

[Hadoop KMS](#) è un server di gestione delle chiavi che fornisce la capacità di implementare servizi crittografici per i cluster Hadoop e può servire come fornitore di chiavi per [Crittografia trasparente in HDFS su Amazon EMR](#). Hadoop KMS in Amazon EMR è installato e abilitato per impostazione predefinita quando si seleziona l'applicazione Hadoop durante l'avvio di un cluster EMR. L'Hadoop KMS non memorizza le chiavi da solo, tranne nel caso di memorizzazione nella cache temporanea. Hadoop KMS funge da proxy tra il fornitore della chiave e il trustee del client verso un keystore di supporto: non è un keystore. Il keystore predefinito che viene creato per Hadoop KMS è Java Cryptography Extension KeyStore (JCEKS). È inclusa anche la policy di forza illimitata JCE, in modo da poter creare chiavi con la lunghezza desiderata. Hadoop KMS supporta anche una gamma di ACL che controllano l'accesso ai tasti e le operazioni chiave indipendentemente da altre applicazioni client come HDFS. La lunghezza predefinita della chiave in Amazon EMR è 256 bit.

Per configurare Hadoop KMS, utilizzare la classificazione `hadoop-kms-site` per modificare le impostazioni. Per configurare ACL, si usa la classificazione `kms-acls`.

Per ulteriori informazioni, consulta la [documentazione di Hadoop KMS](#). Hadoop KMS è utilizzato nella crittografia trasparente HDFS Hadoop. Per ulteriori informazioni su HDFS, consulta l'argomento [HDFS transparent encryption \(Crittografia trasparente HDFS\)](#) nella documentazione di Apache Hadoop.

Note

In Amazon EMR, KMS su HTTPS non è abilitato per impostazione predefinita con Hadoop KMS. Per informazioni su come abilitare KMS su HTTPS, consulta la [documentazione di KMS Hadoop](#).

⚠ Important

Hadoop KMS richiede che i nomi delle chiavi chiave siano minuscoli. Se si utilizza una chiave con caratteri maiuscoli, il cluster fallirà durante l'avvio.

Configurazione di Hadoop KMS in Amazon EMR

Utilizzando Amazon EMR versione 4.6.0 o successive, la `kms-http-port` è 9700 e la `kms-admin-port` è 9701.

È possibile configurare Hadoop KMS al momento della creazione del cluster utilizzando le API di configurazione per i rilasci di Amazon EMR. Di seguito sono riportate le classificazioni degli oggetti di configurazione disponibili per Hadoop KMS:

Classificazioni di configurazione di Hadoop KMS

Classificazione	Nome file
hadoop-kms-site	kms-site.xml
hadoop-kms-acls	kms-acls.xml
hadoop-kms-env	kms-env.sh
hadoop-kms-log4j	kms-log4j.properties

Per impostare ACL di Hadoop KMS utilizzando l'interfaccia a riga di comando (CLI)

- Creare un cluster con Hadoop KMS con ACL utilizzando il seguente comando:

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

Per disabilitare la cache di Hadoop KMS utilizzando l'interfaccia a riga di comando (CLI)

- Creare un cluster con Hadoop KMS `hadoop.kms.cache.enable` impostato su `false` utilizzando il seguente comando:

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

Per impostare le variabili di ambiente nello script **kms-env.sh** utilizzando l'interfaccia a riga di comando (CLI)

- Modificare le impostazioni in **kms-env.sh** tramite la configurazione **hadoop-kms-env**. Creare un cluster con Hadoop KMS utilizzando il seguente comando:

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

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": [
    ]
}
]
}
]

```

Per ulteriori informazioni sulla configurazione di Hadoop KMS, consulta la [documentazione di Hadoop KMS](#).

Crittografia trasparente HDFS su cluster EMR con più nodi master

[Apache Ranger](#) KMS viene utilizzato in un cluster Amazon EMR con più nodi primari per la crittografia trasparente in HDFS.

Apache Ranger KMS archivia la chiave root e le chiavi Encryption Zone (EZ) nel tuo Amazon RDS per un cluster Amazon EMR con più nodi primari. Per abilitare la crittografia trasparente in HDFS su un cluster Amazon EMR con più nodi primari, devi fornire le seguenti configurazioni.

- Amazon RDS o il tuo URL di connessione al server MySQL per archiviare la chiave root Ranger KMS e la chiave EZ
- Nome utente e password per MySQL
- Password per la chiave root Ranger KMS
- File PEM dell'autorità di certificazione (CA) per la connessione SSL al server MySQL

Puoi fornire queste configurazioni utilizzando la classificazione `ranger-kms-dbks-site` e la classificazione `ranger-kms-db-ca`, come indicato nell'esempio seguente.

```

[
  {
    "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"
  }
}
]

```

Di seguito sono riportate le classificazioni degli oggetti di configurazione per Apache Ranger KMS.

Classificazioni di configurazione di Hadoop KMS

Classificazione	Descrizione
ranger-kms-dbks-site	Modifica i valori nel file dbks-site.xml di Ranger KMS.
ranger-kms-site	Modifica i valori nel file ranger-kms-site.xml di Ranger KMS.
ranger-kms-env	Modifica i valori nell'ambiente Ranger KMS.
ranger-kms-log4j	Modifica i valori nel file kms-log4j.properties di Ranger KMS.
ranger-kms-db-ca	Modifica i valori per il file CA su S3 per la connessione SSL di MySQL con Ranger KMS.

Considerazioni

- Si consiglia di crittografare l'istanza Amazon RDS per una maggiore sicurezza. Per ulteriori informazioni, consulta la [Panoramica della crittografia delle risorse Amazon RDS](#).
- Si consiglia di utilizzare database MySQL separati in ogni cluster Amazon EMR con più nodi primari per un maggiore livello di sicurezza.
- Per configurare la crittografia trasparente in HDFS su un cluster Amazon EMR con più nodi primari, devi specificare la classificazione `hdfs-encryption-zones` durante la creazione del

cluster. In caso contrario, Ranger KMS non verrà configurato o avviato. La riconfigurazione della classificazione `hdfs-encryption-zones` o di una qualsiasi classificazione di configurazione Hadoop KMS su un cluster in esecuzione non è supportata su cluster Amazon EMR con più nodi primari.

Creazione ed esecuzione di un'applicazione Hadoop

Argomenti

- [Creazione di file binari con Amazon EMR](#)
- [Elaborazione di dati con lo streaming](#)
- [Elaborazione di dati con un JAR personalizzato](#)

Creazione di file binari con Amazon EMR

Puoi utilizzare Amazon EMR come ambiente di compilazione di programmi da utilizzare nel tuo cluster. I programmi utilizzati con Amazon EMR devono essere compilati su un sistema che esegue la stessa versione di Linux utilizzata da Amazon EMR. Per una versione a 32 bit, la compilazione deve avvenire su un computer a 32 bit o con le opzioni di compilazione incrociata a 32 bit attivate. Per una versione a 64 bit, la compilazione deve avvenire su un computer a 64 bit o con le opzioni di compilazione incrociata a 64 bit attivate. Per ulteriori informazioni sulle versioni di istanze EC2, consulta [Pianificazione e configurazione di istanze EC2](#) nella Guida alla gestione di Amazon EMR. I linguaggi di programmazione supportati includono C++, Python e C#.

La seguente tabella illustra i passaggi necessari per sviluppare e testare la tua applicazione utilizzando Amazon EMR.

Processo per la creazione di un modulo

- 1 Connessione al nodo master del cluster.
- 2 Copia dei file di origine sul nodo master.
- 3 Creazione di file binari con le ottimizzazioni necessarie.
- 4 Copia i file binari dal nodo master su Amazon S3.

Ciascuna di queste fasi è trattata nel dettaglio nelle sezioni che seguono.

Per connettersi al nodo master del cluster

- Segui le istruzioni riportate nella sezione relativa alla [Connessione al nodo principale tramite SSH](#) nella Guida alla gestione di Amazon EMR.

Per copiare i file di origine sul nodo master

1. Inserisci i file di origine in un bucket Amazon S3. Per ulteriori informazioni su come creare i bucket e spostare i dati in Amazon S3, consulta [Guida per l'utente di Amazon Simple Storage Service](#).
2. Creare una cartella sul cluster Hadoop per i file di origine immettendo un comando simile al seguente:

```
mkdir SourceFiles
```

3. Copia i file di origine da Amazon S3 sul nodo master digitando un comando simile al seguente:

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Creazione di file binari con le ottimizzazioni necessarie

La modalità di creazione dei file binari dipende da molti fattori. Segui le istruzioni per i tuoi specifici strumenti di creazione per impostare e configurare il tuo ambiente. Puoi utilizzare i comandi di specifica del sistema di Hadoop per ottenere informazioni sul cluster e determinare come installare il tuo ambiente di compilazione.

Per identificare le specifiche di sistema

- Utilizzare i seguenti comandi per verificare l'architettura utilizzata per creare il file binari.
 - a. Per visualizzare la versione di Debian, immettere il comando seguente:

```
master$ cat /etc/issue
```

L'output è simile al seguente.

```
Debian GNU/Linux 5.0
```

- b. Per visualizzare il nome DNS pubblico e la dimensione del processore, immettere il comando seguente:

```
master$ uname -a
```

L'output è simile al seguente.

```
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. Per visualizzare la velocità del processore, immettere il comando seguente:

```
master$ cat /proc/cpuinfo
```

L'output è simile al seguente.

```
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  
...
```

Una volta creati i file binari, è possibile copiare i file su Amazon S3.

Copia dei file binari dal nodo master su Amazon S3

- Per copiare i file nel bucket Amazon S3, digita il comando seguente:

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

Elaborazione di dati con lo streaming

Lo streaming Hadoop è un'utilità fornita con Hadoop che consente di sviluppare eseguibili MapReduce in linguaggi diversi da Java. Streaming viene implementato sotto forma di un file JAR, perciò è possibile eseguirlo dall'API o riga di comando di Amazon EMR in modo analogo a quanto avviene per un file JAR standard.

Questa sezione descrive come utilizzare Streaming con Amazon EMR.

Note

Apache Hadoop Streaming è uno strumento indipendente. Per questo motivo, tutte le funzioni e i parametri non sono descritti in questa pagina. Per ulteriori informazioni su Hadoop Streaming, consulta la pagina all'indirizzo <http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html>.

Utilizzo dell'utility Hadoop Streaming

Questa sezione descrive come utilizzare per la utility Hadoop Streaming.

Processo di Hadoop

1	<p>Scrivere il mappatore e il riduttore eseguibili nel linguaggio di programmazione preferito.</p> <p>Seguire le istruzioni riportate nella documentazione Hadoop per scrivere i file eseguibili in streaming. I programmi devono leggere i loro dati di ingresso e uscita dai dati di input e uscita output attraverso l'output standard. Per impostazione predefinita, ogni riga di input/output rappresenta un record e la prima scheda di ogni riga viene utilizzata come separatore tra la chiave e il valore.</p>
2	Prova gli eseguibili localmente e caricali su Amazon S3.
3	Utilizza l'interfaccia a riga di comando Amazon EMR o la console di Amazon EMR per eseguire l'applicazione.

Ogni script di mappatura viene avviato come processo separato nel cluster. Ciascun eseguibile del riduttore trasforma l'output dell'eseguibile del mappatore nell'uscita dei dati in base al flusso di lavoro.

I parametri `input`, `output`, `mapper` e `reducer` sono necessari per la maggior parte delle applicazioni di streaming. La seguente tabella descrive questi ed altri parametri opzionali.

Parametro	Descrizione	Obbligatorio
<code>-input</code>	Percorso dei dati di input su Amazon S3.	Sì

Parametro	Descrizione	Obbligatorio
	<p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p> <p>Vincolo: URI. Se non viene specificato alcun protocollo, utilizza il file system predefinito del cluster.</p>	
-output	<p>Percorso su Amazon S3 in cui Amazon EMR carica i dati elaborati.</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p> <p>Vincolo: URI</p> <p>Impostazione predefinita: se il percorso non è specificato, Amazon EMR carica i dati nel percorso specificato da <code>input</code>.</p>	Si
-mapper	<p>Nome del mappatore eseguibile.</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p>	Si
-reducer	<p>Nome del riduttore eseguibile.</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p>	Si

Parametro	Descrizione	Obbligatorio
-cacheFile	<p>Un percorso Amazon S3 contenente file da copiare nella directory di lavoro locale per Hadoop (principalmente per migliorare le prestazioni).</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p> <p>Vincoli: [URI]#[nome symlink da creare nella directory di lavoro]</p>	No
-cacheArchive	<p>File JAR da estrarre nella directory di lavoro</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p> <p>Vincoli: [URI]#[nome directory symlink da creare nella directory di lavoro]</p>	No
-combiner	<p>Combina risultati</p> <p>Tipo: string</p> <p>Impostazione predefinita: nessuna</p> <p>Vincoli: nome della classe Java</p>	No

Il seguente esempio di codice è un eseguibile di mappatura scritto in Python. Questo script fa parte dell'applicazione di esempio WordCount.

```
#!/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)
```

Invio di una fase di streaming

Questa sezione descrive le nozioni di base per l'invio di una fase di streaming a un cluster.

Un'applicazione di streaming legge gli input da un ingresso standard e quindi esegue uno script o un eseguibile (chiamato mappatore) su ogni ingresso. Il risultato di ogni ingresso viene salvato localmente, tipicamente su una partizione Hadoop Distributed File System (HDFS). Dopo che tutti gli input sono stati elaborati dal mappatore, un secondo script o eseguibile (chiamato riduttore) elabora i risultati del mappatore. I risultati del riduttore vengono inviati all'output standard. È possibile concatenare una serie di fasi di Streaming, in cui l'output di una fase diventa l'input di un'altra fase.

Il mappatore e il riduttore possono essere referenziati come file oppure si può fornire una classe Java. È possibile implementare il mappatore e il riduttore in qualsiasi lingua supportata, inclusi Ruby, Perl, Python, PHP o Bash.

Invio di una fase di streaming utilizzando la console

Questo esempio descrive come utilizzare la console di Amazon EMR per inviare una fase di Streaming personalizzata a un cluster in esecuzione.

Invio di una fase di streaming

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. In Cluster List (Elenco cluster), selezionare il nome del cluster.
3. Scorrere fino alla sezione Steps (Fasi) ed espanderla, quindi scegliere Add step (Aggiungi fase).
4. Nella finestra di dialogo Add Step (Aggiungi fase):
 - Per Step type (Tipo di fase), scegliere Streaming program (Programma Streaming).
 - Per Name (Nome), accettare il nome di default, ovvero Streaming program (Programma Streaming), o digitare un nuovo nome.

- Per Mapper (Mappatore), digitare o aprire la posizione della classe del mapper in Hadoop, o un bucket S3 in cui risiede l'eseguibile del mappatore, come un programma Python. Il valore del percorso deve essere nel formato *BucketName/path/MappperExecutable*.
 - Per Reducer (Riduttore), digitare o aprire la posizione della classe del riduttore in Hadoop, o un bucket S3 in cui risiede l'eseguibile del riduttore, come un programma Python. Il valore del percorso deve essere nel formato *BucketName/path/MappperExecutable*. Amazon EMR supporta la parola chiave speciale aggregata. Per ulteriori informazioni, vedere la libreria di aggregazione fornita da Hadoop.
 - Per Input S3 location (Percorso input S3), digitare o aprire il percorso dei dati di input.
 - Per Output S3 location (Percorso S3 output), digita il nome del bucket di output Amazon S3.
 - Lasciare vuoto il campo Arguments (Argomenti).
 - Per Action on failure (Operazione in caso di errore), accettare l'opzione predefinita, ovvero Continue (Continua).
5. Scegliere Add (Aggiungi). La fase viene visualizzata nella console con lo stato Pending (In attesa).
 6. Durante l'esecuzione della fase, lo stato passa da Pending (In attesa) a Running (In esecuzione) a Completed (Completata). Per aggiornare lo stato, scegliere l'icona Refresh (Aggiorna) sopra la colonna Actions (Operazioni).

AWS CLI

Questi esempi illustrano come utilizzare la AWS CLI per creare un cluster e inviare una fase di Streaming.

Creazione di un cluster e invio di una fase di streaming utilizzando la AWS CLI

- Per creare un cluster e inviare una fase di streaming utilizzando AWS CLI digitare il comando seguente e sostituire *myKey* con il nome della coppia di chiavi di EC2. Nota che l'argomento per `--files` dovrebbe essere il percorso Amazon S3 della posizione dello script e gli argomenti per `-mapper` e `-reducer` dovrebbero essere i nomi dei rispettivi file di script.

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 --
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

I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo master e le istanze rimanenti vengono avviate come nodi principali. Tutti i nodi utilizzano il tipo di istanza specificato nel comando.

Note

Se in precedenza non sono stati creati il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Elaborazione di dati con un JAR personalizzato

Un JAR personalizzato esegue un programma Java compilato che viene caricato su Amazon S3. Devi compilare il programma in base alla versione di Hadoop che desideri avviare e invia una fase `CUSTOM_JAR` al tuo cluster Amazon EMR. Per ulteriori informazioni sulla compilazione di un file JAR, consulta [Creazione di file binari con Amazon EMR](#).

Per ulteriori informazioni sulla creazione di un'applicazione Hadoop MapReduce, consulta il [Tutorial MapReduce](#) nella documentazione di Apache Hadoop.

Argomenti

- [Invio di una fase JAR personalizzata](#)

Invio di una fase JAR personalizzata

Un JAR personalizzato esegue un programma Java compilato che viene caricato su Amazon S3. Devi compilare il programma in base alla versione di Hadoop che desideri avviare e invia una fase CUSTOM_JAR al tuo cluster Amazon EMR. Per ulteriori informazioni sulla compilazione di un file JAR, consulta [Creazione di file binari con Amazon EMR](#).

Per ulteriori informazioni sulla creazione di un'applicazione Hadoop MapReduce, consulta il [Tutorial MapReduce](#) nella documentazione di Apache Hadoop.

Questa sezione descrive le basi per l'invio di una fase JAR personalizzata in Amazon EMR. L'invio di una fase JAR personalizzata consente di scrivere uno script per elaborare i dati con il linguaggio di programmazione Java.

Invio di una fase JAR personalizzata utilizzando la console

Questo esempio descrive come utilizzare la console di Amazon EMR per inviare una fase JAR personalizzata a un cluster in esecuzione.

Per inviare una fase JAR personalizzata utilizzando la console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. In Cluster List (Elenco cluster), selezionare il nome del cluster.
3. Scorrere fino alla sezione Steps (Fasi) ed espanderla, quindi scegliere Add step (Aggiungi fase).
4. Nella finestra di dialogo Add Step (Aggiungi fase):
 - Per Step Type (Tipo di fase) selezionare Custom JAR (JAR personalizzato).
 - Per Name (Nome), accettare il nome di default, ovvero Custom JAR (JAR personalizzato), o digitare un nuovo nome.
 - Per JAR S3 location (Posizione JAR S3), digitare o selezionare il percorso del file JAR. La posizione JAR può essere un percorso in S3 o una classe java completa nel classpath.
 - Per Arguments (Argomenti), digitare eventuali argomenti obbligatori come stringhe separate da spazi o lasciare il campo vuoto.
 - Per Action on failure (Operazione in caso di errore), accettare l'opzione predefinita, ovvero Continue (Continua).
5. Scegliere Add (Aggiungi). La fase viene visualizzata nella console con lo stato Pending (In attesa).

6. Durante l'esecuzione della fase, lo stato passa da Pending (In attesa) a Running (In esecuzione) a Completed (Completata). Per aggiornare lo stato, scegliere l'icona Refresh (Aggiorna) sopra la colonna Actions (Operazioni).

Avvio di un cluster e invio di una fase JAR personalizzata con l'AWS CLI

Per avviare un cluster e inviare una fase JAR personalizzata con l'AWS CLI

Per avviare un cluster e inviare una fase JAR personalizzata con l'AWS CLI, digita il sottocomando `create-cluster` con il parametro `--steps`.

- Per avviare un cluster e inviare una fase JAR personalizzata, digitare il comando seguente, sostituire *myKey* con il nome della coppia di chiavi EC2 e sostituire *mybucket* con il nome del bucket.

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \  
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \  
--steps Type=CUSTOM_JAR,Name="Custom JAR  
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket"]
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

Quando specifichi il numero di istanze senza il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi principali. Tutti i nodi utilizzano il tipo di istanza specificato nel comando.

Note

Se in precedenza non sono stati creati il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Dipendenze da terze parti

A volte può essere necessario includere nel percorso della classe di MapReduce JAR per l'uso con il programma. Ci sono due possibilità per farlo:

- Includere l'`--libjars s3://URI_to_JAR` nelle opzioni di fase per la procedura in [Avvio di un cluster e invio di una fase JAR personalizzata con l'AWS CLI](#).
- Avvia il cluster con un'impostazione `mapreduce.application.classpath` modificata in `mapred-site.xml`. Utilizza la classificazione della configurazione `mapred-site`. Il cluster creato con la fase utilizzando la AWS CLI dovrebbe assomigliare a quanto segue:

```
aws emr create-cluster --release-label emr-5.36.1 \
--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"
    }
  }
]
```

L'elenco dei percorsi separati da virgole dovrebbe essere aggiunto al percorso della classe JVM per ogni attività.

Attiva il riconoscimento non uniforme degli accessi alla memoria per i container YARN

Con le versioni 6.x e successive di Amazon EMR, puoi utilizzare l'accesso non uniforme alla memoria (NUMA) per l'elaborazione multipla dei dati sui cluster. NUMA è un modello di progettazione della memoria del computer in cui il processore può accedere alla propria memoria locale più velocemente che della memoria di un altro processore o da quella condivisa tra processori. I container YARN hanno prestazioni migliori con NUMA perché possono collegarsi a uno specifico nodo NUMA che serve tutte le successive allocazioni di memoria. Ciò riduce il numero di volte in cui il cluster deve accedere alla memoria remota.

Puoi attivare il supporto NUMA per il container YARN quando la macchina con nodi worker è un nodo multi-NUMA. Per confermare se un nodo worker è un nodo a NUMA singolo o multi-NUMA, esegui il comando che segue.

```
lscpu | grep -i numa
NUMA node(s): 2
```

In generale, le istanze più grandi di 12x hanno due nodi NUMA. Questo non si applica ai istanze metal.

Per attivare il riconoscimento di NUMA per i container YARN

1. Utilizza la seguente configurazione `yarn-site` nel tuo cluster Amazon EMR 6.x.

```
[
  {
    "classification":"yarn-site",
    "properties":{
      "yarn.nodemanager.linux-container-executor.nonsecure-mode.local-
user":"yarn",
      "yarn.nodemanager.linux-container-executor.group":"yarn",
      "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor",
      "yarn.nodemanager.numa-awareness.enabled":"true",
      "yarn.nodemanager.numa-awareness.numactl.cmd":"/usr/bin/numactl",
      "yarn.nodemanager.numa-awareness.read-topology":"true"
    },
    "configurations":[]
  }
]
```

```
] ]
```

2. Fornisci la seguente azione di bootstrap nel tuo cluster.

```
#!/bin/bash

sudo yum -y install numactl
echo 1 | sudo tee /proc/sys/kernel/numa_balancing

echo "banned.users=mapred,bin,hdfs" >> /etc/hadoop/conf/container-executor.cfg
rm -rf /var/log/hadoop-yarn/
sudo chown -R yarn:hadoop /var/log/hadoop-yarn/
sudo chmod 755 -R /var/log/hadoop-yarn/

sudo chmod 6050 /etc/hadoop/conf/container-executor.cfg

mkdir /mnt/yarn && sudo chmod 755 -R /mnt/yarn && sudo chown -R yarn:hadoop /mnt/
yarn
mkdir /mnt1/yarn && sudo chmod 755 -R /mnt1/yarn && sudo chown -R yarn:hadoop /
mnt1/yarn
mkdir /mnt2/yarn && sudo chmod 755 -R /mnt2/yarn && sudo chown -R yarn:hadoop /
mnt2/yarn
```

3. Ogni container deve essere a conoscenza di NUMA. Puoi segnalare la macchina virtuale Java (JVM) in ogni container con un flag NUMA. Ad esempio, per notificare alla JVM di utilizzare NUMA in un processo MapReduce, aggiungi le seguenti proprietà in `mapred-site.xml`.

```
<property>
  <name>mapreduce.reduce.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
<property>
  <name>mapreduce.map.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
```

4. Per verificare di aver attivato NUMA, cerca uno dei file di log di NodeManager con il seguente comando.

```
grep "NUMA resources allocation is enabled," *
```

Per verificare che NodeManager abbia assegnato le risorse del nodo NUMA a un container, cerca nel log NodeManager con il seguente comando, sostituendo `<container_id>` con il tuo ID container.

```
grep "NUMA node" | grep <container_id>
```

Cronologia delle versioni di Hadoop

Nelle tabelle seguenti sono elencate la versione di Hadoop inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Hadoop

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
		mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager
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-resource-manager
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-resource-manager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager
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-resource-manager
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-resource-manager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager
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-resource-manager
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-resource-manager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.0	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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.8.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.4	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.2	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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.0	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

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
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-resource-manager
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-resource-manager
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-resource-manager

Etichetta di rilascio di Amazon EMR	Versione Hadoop	Componenti installati con Hadoop
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
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

Note sul rilascio di Hadoop per versione

[Amazon EMR 6.6.0 - Note sul rilascio di Hadoop](#)

Amazon EMR 6.6.0 - Note sul rilascio di Hadoop

Amazon EMR 6.6.0 - Modifiche a Hadoop

Tipo	Descrizione
Bug	Corretti i record duplicati durante la lettura di file di testo BZip2.
Backport	HADOOP-18136 : verifica della gestione <code>FileUtils.unTar()</code> dei file .tar mancanti
Backport	HADOOP-17627 : esecuzione del backport a branch-3.2 HADOOP-17371, HADOOP-17621, HADOOP-17625 per aggiornare Jetty a 9.4.39
Backport	HADOOP-17655 : aggiornamento di Jetty a 9.4.40
Backport	HADOOP-17796 : aggiornamento della versione di Jetty a 9.4.43
Backport	HADOOP-17661 : mancata esecuzione del parsing di pom.xml da parte di mvn versions:set
Backport	HADOOP-17236 : passaggio di snakeyaml a 1.26 per mitigare CVE-2017-18640
Backport	HADOOP-16717 : rimozione della dipendenza a isLog4jLogger di GenericsUtil su Log4JLoggerAdapter
Backport	HADOOP-17633 : passaggio di json-smart a 2.4.2 e nimbus-jose-jwt a 9.8 a causa di CVE

Tipo	Descrizione
Backport	HADOOP-17844 : aggiornamento di JSON smart a 2.4.7
Backport	HADOOP-17972 : esecuzione del backport di HADOOP-17683 (aggiornato commons-io a 2.8.0) per branch-3.2
Backport	HADOOP-16555 : aggiornamento di commons-compress alla 1.19
Backport	HADOOP-17370 : aggiornamento di commons-compress a 1.21
Backport	HADOOP-17096 : correzione dell'offset del buffer di ingresso ZStandardCompressor
Backport	HADOOP-17112 : mancata ammissione degli spazi bianchi nei percorsi durante il salvataggio dei file su s3a tramite committer
Backport	HADOOP-13500 : Sincronizzazione dell'iterazione dell'oggetto delle proprietà di configurazione
Backport	HDFS-14099 : mancato riconoscimento del descrittore di frame durante la decompressione di più fotogrammi in ZStandardDecompressor
Backport	HDFS-16410 : analisi Xml non sicura in OfflineEditSXMLLoader
Backport	HDFS-14498 : possibilità di esecuzione di un ciclo infinito sul file per il quale la creazione non è riuscita da parte di LeaseManager
Backport	HDFS-15290 : NPE in HttpServer durante l'avvio di NameNode

Tipo	Descrizione
Backport	HDFS-15293 : attenuazione della condizione per accettare un fsimage quando si riceve un checkpoint
Backport	HDFS-12979 : StandbyNode dovrebbe caricare FSImage su ObserverNode dopo il checkpoint
Backport	YARN-10538 : aggiunta di nodi di riattivazione all'elenco dei nodi aggiornati restituiti all'AM
Backport	YARN-10472 : esecuzione del backport di YARN-10314 (YarnClient avvia NoClassDefFoundError per WebSocketException con solo i jar client schermati) a branch-3.2
Backport	YARN-9968 : Public Localizer termina in NodeManager a causa di NullPointerException
Backport	YARN-10651 : CapacityScheduler si è bloccato con NPE in AbstractYarnScheduler.updateNodeResource()
Backport	YARN-9339 : parametro delle app in sospenso non corretto dopo lo spostamento dell'app in una nuova coda
Backport	YARN-10438 : gestisce i valori null containerId in ClientRMService#getContainerReport()
Backport	YARN-7266 : ATS 1.5 non si avvia se i file RollingLevelDb sono danneggiati o mancanti

Tipo	Descrizione
Backport	YARN-9063 : ATS 1.5 non si avvia se i file <code>RollingLevelDb</code> sono danneggiati o mancanti
Backport	YARN-9848 : ripristino di YARN-4946 (RM non deve considerare un'applicazione come <code>COMPLETED</code> quando l'aggregazione del log non è in uno stato terminale).

Apache HBase

[HBase](#) è un database distribuito, non relazionale, open source sviluppato come parte del progetto Hadoop di Apache Software Foundation. HBase viene eseguito nel File system distribuito Hadoop (HDFS) per offrire funzionalità di database non relazionale per l'ecosistema Hadoop. HBase è incluso con Amazon EMR versione 4.6.0 e successive.

HBase funziona perfettamente con Hadoop, condividendone il file system e agendo come ingresso e uscita diretti al framework e motore di esecuzione MapReduce. HBase si integra anche con Apache Hive, abilitando query di tipo SQL su tabelle HBase, esegue l'unione con tabelle basate su Hive e il supporto per Java Database Connectivity (JDBC). Per ulteriori informazioni su HBase, consulta le sezioni relative ad [Apache HBase](#) e alla [documentazione HBase](#) nel sito Web di Apache. Per un esempio di come utilizzare HBase con Hive, consulta l'articolo pubblicato sul Blog sui Big Data AWS [Combine NoSQL and Massively Parallel Analytics Using Apache HBase and Apache Hive on Amazon EMR \(Combinazione di NoSQL e analisi dei dati MPP \(Massively Parallel Processing\) utilizzando Apache HBase e Apache Hive su Amazon EMR\)](#).

Con HBase su Amazon EMR puoi anche eseguire il backup dei dati HBase direttamente in Amazon Simple Storage Service (Amazon S3) e il ripristino da un backup creato in precedenza all'avvio di un cluster HBase. Amazon EMR offre opzioni aggiuntive per l'integrazione con Amazon S3 per la persistenza dei dati e il ripristino di emergenza.

- HBase su Amazon S3: con Amazon EMR versione 5.2.0 e successive, puoi utilizzare HBase su Amazon S3 per archiviare la directory principale e i metadati HBase del cluster direttamente in Amazon S3. In seguito puoi avviare un nuovo cluster, indirizzandolo alla posizione della directory principale in Amazon S3. Il percorso HBase in Amazon S3 può essere utilizzato da un solo cluster alla volta, a eccezione di una replica di lettura al cluster. Per ulteriori informazioni, consulta [HBase su Amazon S3 \(modalità di archiviazione Amazon S3\)](#).
- Repliche di lettura ad HBase: Amazon EMR versione 5.7.0 e successive con HBase su Amazon S3 supporta la replica di lettura ai cluster. Una replica di lettura al cluster fornisce accesso in sola lettura ai file e ai metadati dello store di un cluster principale per operazioni di sola lettura. Per ulteriori informazioni, consulta [Utilizzo della replica di lettura al cluster](#).
- Snapshot HBase - In alternativa a HBase su Amazon S3, con EMR versione 4.0 e successive puoi creare snapshot dei dati HBase direttamente in Amazon S3 e, di conseguenza, recuperare dati utilizzando gli snapshot. Per ulteriori informazioni, consulta [Utilizzo di snapshot HBase](#).

⚠ Important

Per il dimensionamento dei cluster HBase di Amazon EMR, non consigliamo di utilizzare il [dimensionamento gestito](#) o il [dimensionamento con policy personalizzate](#) con i cluster HBase.

La seguente tabella indica la versione di HBase inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con HBase.

Per la versione dei componenti installati con HBase in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di HBase per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-6.14.0	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-ftpfs-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 è uno strumento operativo separato per la riparazione di aree HBase e tabelle di sistema. In Amazon EMR versione 6.1.0 e successive, `hbase-hbck2.jar` è fornito in `/usr/lib/hbase-operator-tools/` sul nodo master. Per ulteriori informazioni su come creare e utilizzare lo strumento, vedere [HBase HBCK2](#).

La seguente tabella indica la versione di HBase inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con HBase.

Per la versione dei componenti installati con HBase in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di HBase per `emr-5.36.1`

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
<code>emr-5.36.1</code>	HBase 1.4.13	<code>emrfs</code> , <code>emr-ddb</code> , <code>emr-goodies</code> , <code>emr-kinesis</code> , <code>emr-s3-distcp</code> , <code>hadoop-client</code> , <code>hadoop-hdfs-datanode</code> , <code>hadoop-hdfs-library</code> , <code>hadoop-hdfs-namenode</code> , <code>hadoop-https-server</code> , <code>hadoop-kms-server</code> , <code>hadoop-mapred</code> , <code>hadoop-yarn-nodemanager</code> , <code>hadoop-yarn-resource-manager</code> , <code>hadoop-yarn-timeline-server</code> , <code>hbase-master</code> , <code>hbase-client</code> , <code>hbase-region-server</code> , <code>hbase-rest-server</code> , <code>hbase-thrift-server</code> , <code>zookeeper-client</code> , <code>zookeeper-server</code>

Argomenti

- [Creazione di un cluster con HBase](#)

- [HBase su Amazon S3 \(modalità di archiviazione Amazon S3\)](#)
- [Utilizzo della shell HBase](#)
- [Accesso a tabelle HBase con Hive](#)
- [Utilizzo di snapshot HBase](#)
- [Configurazione di HBase](#)
- [Visualizzazione dell'interfaccia utente di HBase](#)
- [Visualizzazione di file di log HBase](#)
- [Monitoraggio di HBase con Ganglia](#)
- [Migrazione da versioni HBase precedenti](#)
- [Cronologia dei rilasci di HBase](#)

Creazione di un cluster con HBase

Le procedure in questa sezione illustrano le nozioni di base per l'avvio di un cluster utilizzando AWS Management Console e AWS CLI. Per informazioni dettagliate su come pianificare, configurare e avviare cluster EMR, consulta la sezione relativa alla [Pianificazione e configurazione dei cluster](#) nella Guida alla gestione di Amazon EMR.

Creazione di un cluster con HBase utilizzando la console

Per le fasi rapide di avvio di cluster con la console, consulta [Nozioni di base su Amazon EMR](#) nella Guida alla gestione di Amazon EMR.

Per avviare un cluster con HBase installato utilizzando la console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Seleziona Create cluster (Crea cluster) e Go to advanced options (Vai alle opzioni avanzate).
3. Per Software Configuration (Configurazione del software), scegli una versione Amazon 4.6.0 o successive (raccomandiamo la versione più recente). Scegli HBase e altre applicazioni come desiderato.
4. Con Amazon EMR versione 5.2.0 e successive, in HBase Storage Settings (Impostazioni di archiviazione HBase), seleziona HDFS o S3. Per ulteriori informazioni, consulta [HBase su Amazon S3 \(modalità di archiviazione Amazon S3\)](#).
5. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Creazione di un cluster con HBase utilizzando la AWS CLI

Utilizza il comando seguente per creare un cluster con HBase installato:

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Se utilizzi HBase su Amazon S3, specifica l'opzione `--configurations` con un riferimento a un oggetto di configurazione JSON. L'oggetto di configurazione deve contenere una classificazione `hbase-site` che specifica il percorso in Amazon S3 in cui i dati HBase vengono archiviati utilizzando la proprietà `hbase.rootdir`. Deve inoltre contenere una classificazione `hbase`, che specifica `s3` usando la proprietà `hbase.emr.storageMode`. Nell'esempio seguente viene mostrato uno snippet JSON con queste impostazioni di configurazione.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3"  
    }  
  }  
]
```

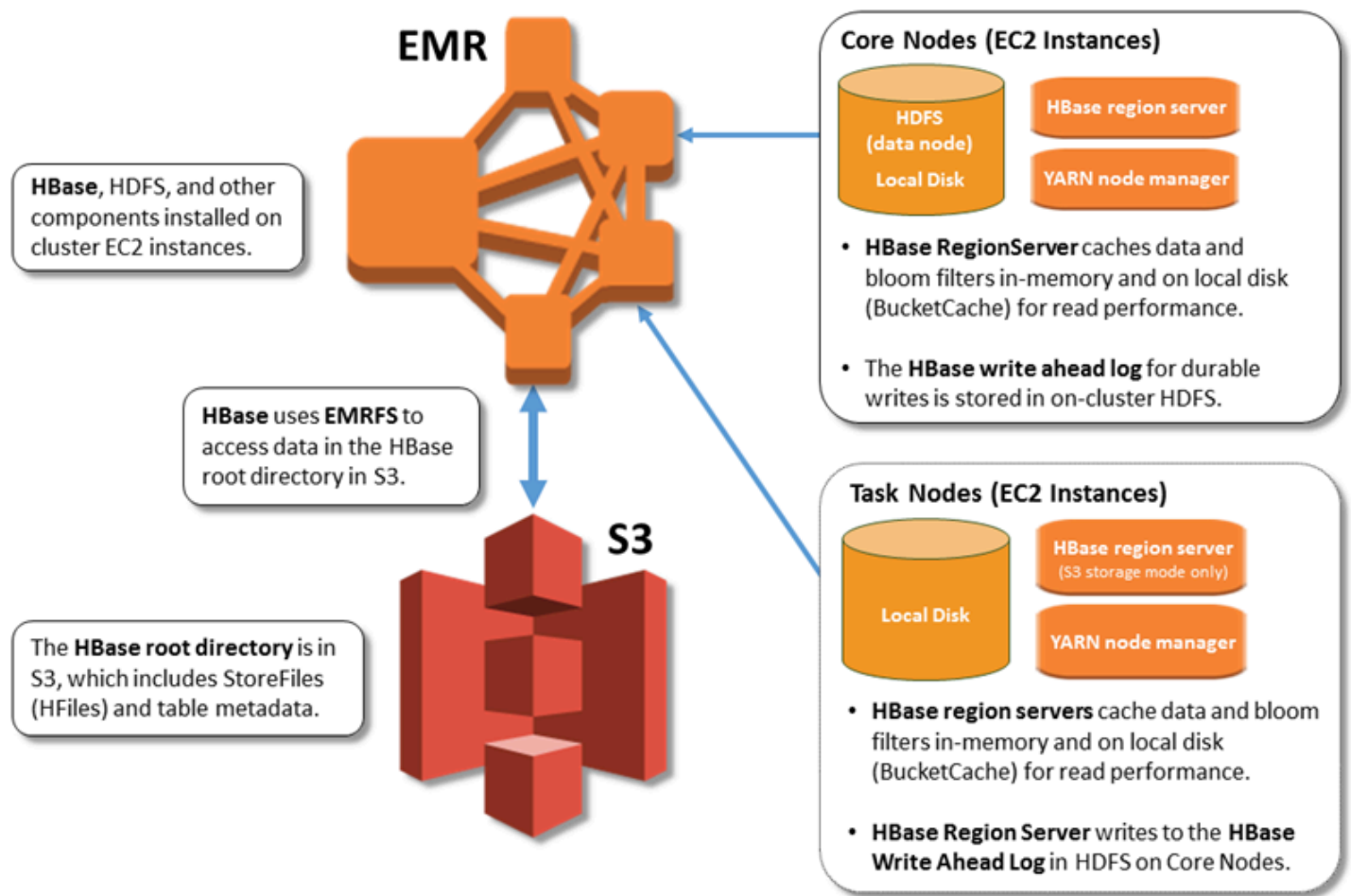
Per ulteriori informazioni su HBase su Amazon S3, consulta [HBase su Amazon S3 \(modalità di archiviazione Amazon S3\)](#). Per ulteriori informazioni sulle classificazioni, consulta [Configurazione delle applicazioni](#).

HBase su Amazon S3 (modalità di archiviazione Amazon S3)

Quando esegui HBase su Amazon EMR versione 5.2.0 o successive, puoi abilitare HBase su Amazon S3, che offre i seguenti vantaggi:

- La directory principale HBase viene archiviata in Amazon S3, inclusi i file dello store HBase e i metadati delle tabelle. Questi dati sono persistenti al di fuori del cluster, sono disponibili tra zone di disponibilità Amazon EC2 e non è necessario eseguire il ripristino utilizzando snapshot o altri metodi.
- Con i file dello store in Amazon S3 puoi dimensionare il cluster Amazon EMR per i tuoi requisiti di calcolo anziché i requisiti dei dati, con replica 3x in HDFS.
- Utilizzando Amazon EMR versione 5.7.0 o successive, puoi impostare una replica di lettura al cluster, che consente di mantenere copie di sola lettura dei dati in Amazon S3. Puoi accedere ai dati dalla replica di lettura al cluster per eseguire operazioni di lettura simultaneamente e nel caso in cui il cluster principale diventi indisponibile.
- In Amazon EMR versione 6.2.0 e successive, il tracciamento HFile persistente utilizza una tabella di sistema HBase chiamata `hbase:storefile` per tracciare direttamente i percorsi HFile utilizzati per le operazioni di lettura. Questa caratteristica è abilitata per impostazione predefinita e non richiede la migrazione manuale.

La figura seguente mostra i componenti di HBase rilevanti per HBase su Amazon S3.



Abilitare HBase su Amazon S3

Puoi abilitare HBase su Amazon S3 utilizzando la console di Amazon EMR, la AWS CLI o l'API di Amazon EMR. La configurazione è un'opzione durante la creazione del cluster. Quando utilizzi la console, puoi scegliere l'impostazione utilizzando **Advanced options** (Opzioni avanzate). Quando utilizzi AWS CLI, utilizza l'opzione `--configurations` per fornire un oggetto di configurazione JSON. Le proprietà dell'oggetto di configurazione specificano la modalità di archiviazione e il percorso della directory principale in Amazon S3. Il percorso Amazon S3 specificato deve trovarsi nella stessa regione del cluster Amazon EMR. Solo un cluster attivo alla volta può utilizzare la stessa directory principale HBase in Amazon S3. Per le fasi della console e un esempio di creazione cluster dettagliato utilizzando AWS CLI, consulta [Creazione di un cluster con HBase](#). Un esempio di oggetto di configurazione è mostrato nel seguente snippet JSON.

```
{
  "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3"
  }
}
```

Note

Se utilizzi un bucket Amazon S3 come `rootdir` per HBase, dovrai aggiungere una barra alla fine dell'URI Amazon S3. Per evitare problemi occorre utilizzare, ad esempio, `"hbase.rootdir: s3://my-bucket/"` anziché `"hbase.rootdir: s3://my-bucket"`.

Utilizzo della replica di lettura al cluster

Dopo aver configurato un cluster principale utilizzando HBase su Amazon S3, puoi creare e configurare una replica di lettura al cluster che fornisce accesso in sola lettura agli stessi dati del cluster principale. Ciò è utile quando è necessario accesso simultaneo per eseguire query sui dati o accesso ininterrotto se il cluster principale diventa non disponibile. La funzione replica di lettura è disponibile con Amazon EMR versione 5.7.0 e successive.

Il cluster principale e la replica di lettura al cluster vengono impostati nello stesso modo con una differenza importante. Entrambi fanno riferimento allo stesso percorso `hbase.rootdir`. Tuttavia, la classificazione `hbase` per la replica di lettura al cluster include la proprietà `"hbase.emr.readreplica.enabled": "true"`.

Ad esempio, data la classificazione JSON per il cluster principale come mostrato in precedenza nell'argomento, la configurazione di una replica di lettura al cluster è la seguente:

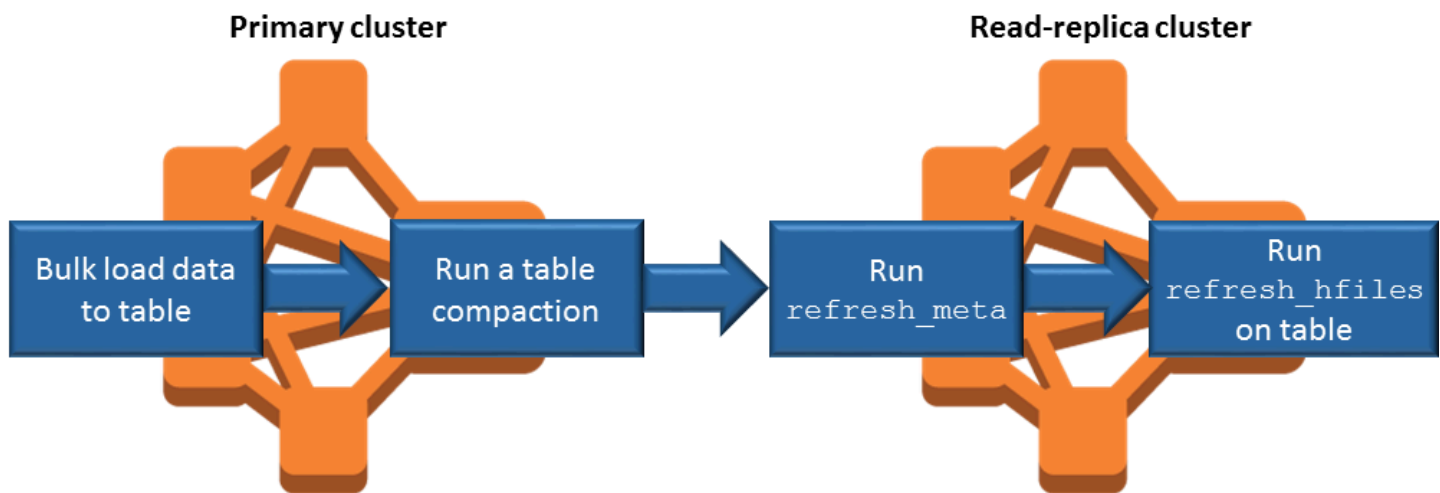
```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
```

```
},  
{  
  "Classification": "hbase",  
  "Properties": {  
    "hbase.emr.storageMode": "s3",  
    "hbase.emr.readreplica.enabled": "true"  
  }  
}
```

Sincronizzazione della replica di lettura quando si aggiungono dati

Poiché la replica di lettura utilizza file e metadati dello store HBase che il cluster principale scrive in Amazon S3, la replica di lettura è attuale tanto quanto il datastore Amazon S3. Le seguenti indicazioni consentono di ridurre il ritardo tra il cluster principale e la replica di lettura durante la scrittura dei dati.

- Se possibile, caricare i dati in blocco nel cluster principale. Per ulteriori informazioni, consulta la sezione relativa al [Bulk loading \(Caricamento in blocco\)](#) nella documentazione di Apache HBase.
- Dopo che i dati vengono aggiunti, i file dello store devono essere scritti al più presto in Amazon S3 mediante uno svuotamento. Eseguire lo svuotamento manualmente o ottimizzare le impostazioni di svuotamento per ridurre il ritardo.
- Se è possibile che vengano eseguite compattazioni automatiche, eseguire una compattazione manuale per evitare incoerenze quando le compattazioni vengono attivate.
- Nella replica di lettura al cluster, quando sono state apportate modifiche ai metadati, ad esempio quando si verificano divisioni regione HBase o compattazioni o quando vengono aggiunte o rimosse tabelle, eseguire il comando `refresh_meta`.
- Nella replica di lettura al cluster, eseguire il comando `refresh_hfiles` quando si aggiungono o modificano record in una tabella.



Tracciamento HFile persistente

Il tracciamento HFile persistente utilizza una tabella di sistema HBase chiamata `hbase:storefile` per tracciare direttamente i percorsi HFile utilizzati per le operazioni di lettura. I nuovi percorsi HFile vengono aggiunti alla tabella come i dati aggiuntivi vengono aggiunti ad HBase. Questo rimuove le operazioni di rinomina come meccanismo di commit nelle operazioni HBase del percorso di scrittura critico e ottimizza il tempo di ripristino quando si apre una regione HBase leggendo dalla tabella di sistema `hbase:storefile` anziché dall'elenco delle directory del file system. Questa funzione è abilitata per impostazione predefinita su Amazon EMR 6.2.0 e versioni successive e non richiede alcuna procedura di migrazione manuale.

Note

Il tracciamento HFile persistente utilizzando la tabella di sistema Storefile di HBase non supporta la funzionalità di replica della regione HBase. Per ulteriori informazioni sulla replica della regione HBase, consulta [Lecture ad alta disponibilità coerenti con la tempistica](#).

Disabilitare il tracciamento HFile persistente

Il tracciamento HFile persistente è abilitato per impostazione predefinita a partire da EMR rilascio 6.2.0. Per disabilitare il tracciamento HFile persistente, specifica la seguente override di configurazione durante l'avvio di un cluster:

```
{
  "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.storefile.tracking.persist.enabled":"false",

"hbase.hstore.engine.class":"org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
}
}
```

Note

Durante la riconfigurazione del cluster Amazon EMR, tutti i gruppi di istanze devono essere aggiornati.

Sincronizzazione manuale della tabella Storefile

La tabella Storefile viene aggiornata man mano che vengono creati nuovi HFile. Tuttavia, se la tabella Storefile non è sincronizzata con i file di dati per qualsiasi motivo, è possibile utilizzare i seguenti comandi per sincronizzare manualmente i dati:

Sincronizza la tabella Storefile in una regione online:

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

Sincronizza la tabella Storefile in una regione offline:

- Rimuove lo znode della tabella Storefile.

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- Assegna la regione (da eseguire in "hbase shell").

```
hbase cli> assign '<region name>'
```

- Se l'assegnazione ha esito negativo.

```
hbase cli> disable '<table name>'
```



```
hbase cli> enable '<table name>'
```

Ridimensionamento della tabella Storefile

La tabella Storefile è suddivisa in quattro regioni per impostazione predefinita. Se la tabella Storefile presenta ancora un'ingente mole di scrittura, può essere ulteriormente divisa in modalità manuale.

Per dividere una regione critica specifica, utilizza il seguente comando (da eseguire in "hbase shell").

```
hbase cli> split '<region name>'
```

Per dividere la tabella, utilizza il seguente comando (da eseguire in "hbase shell").

```
hbase cli> split 'hbase:storefile'
```

Considerazioni operative

I server della regione HBase utilizzano BlockCache per archiviare letture di dati in memoria e BucketCache per archiviare letture di dati su disco locale. Inoltre, i server della regione utilizzano MemStore per archiviare scritture di dati in memoria e log write-ahead per archiviare scritture di dati in HDFS prima che i dati vengano scritti in file dello store HBase in Amazon S3. Le prestazioni di lettura del cluster fanno riferimento alla frequenza con cui un record può essere recuperato dalle cache in memoria o su disco. Un mancato riscontro nella cache comporta che il record venga letto da StoreFile in Amazon S3, con una latenza notevolmente più elevata e una deviazione standard più alta rispetto alla lettura da HDFS. Inoltre, le frequenze di richiesta massime per Amazon S3 sono inferiori rispetto a quelle che possono essere ottenute dalla cache locale, pertanto il caching dei dati può essere importante per i carichi di lavoro gravosi in lettura. Per maggiori informazioni sulle prestazioni di Amazon S3, consulta la sezione [Ottimizzazione delle prestazioni](#) nella Guida per l'utente di Amazon Simple Storage Service.

Per migliorare le prestazioni, ti consigliamo di memorizzare nella cache la massima quantità di set di dati possibile nello storage dell'istanza EC2. Poiché BucketCache utilizza l'archiviazione dell'istanza EC2 del server della regione, puoi scegliere un tipo di istanza EC2 con un archivio istanza sufficiente e aggiungere archiviazione ad Amazon EBS per tenere conto delle dimensioni della cache richieste. Puoi anche incrementare le dimensioni BucketCache sulle instance store e i volumi EBS collegati utilizzando la proprietà `hbase.bucketcache.size`. L'impostazione predefinita è 8192 MB.

Per le operazioni di scrittura, la frequenza degli svuotamenti MemStore e il numero di StoreFile presenti durante compattamenti minori e maggiori possono contribuire in modo significativo a un incremento dei tempi di risposta del server della regione. Per prestazioni ottimali, valutare se incrementare le dimensioni dello svuotamento MemStore e del moltiplicatore blocco HRegion, con conseguente aumento del tempo trascorso tra compattamenti maggiori, ma anche del ritardo in consistenza se si utilizza una replica di lettura. In alcuni casi, è possibile migliorare le prestazioni utilizzando dimensioni dei blocchi di file più grandi (ma inferiori a 5 GB) per attivare la funzionalità di caricamento in più parti Amazon S3 in EMRFS. La dimensione predefinita del blocco di Amazon EMR è 128 MB. Per ulteriori informazioni, consulta [Configurazione HDFS](#). Raramente vediamo clienti che superano le dimensioni dei blocchi da 1 GB mentre eseguono il benchmarking delle prestazioni con scarichi e compattazioni. Inoltre, compattazioni HBase e server della regione garantiscono prestazioni ottimali quando il numero di StoreFile da compattare è inferiore.

Il rilascio delle tabelle su Amazon S3 può richiedere molto tempo perché è necessario rinominare directory di grandi dimensioni. Valuta se disabilitare le tabelle anziché rilasciarle.

Esiste un processo di pulizia HBase che rimuove vecchi file WAL e file dello store. Con Amazon EMR versione 5.17.0 e successive, il processo di eliminazione è abilitato a livello globale e le seguenti proprietà di configurazione possono essere utilizzate per controllare il comportamento di tale processo.

Proprietà di configurazione	Valore predefinito	Descrizione
<code>hbase.regionserver.hfilecleaner.large.thread.count</code>	1	Il numero di thread allocato per pulire HFile di grandi dimensioni scaduti.
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	Il numero di thread allocato per pulire HFile di piccole dimensioni scaduti.
<code>hbase.cleaner.scan.dir.concurrent.size</code>	Impostato su un quarto di tutti i core disponibili.	Il numero di thread per eseguire una scansione delle directory oldWALs.
	2	

Proprietà di configurazione	Valore predefinito	Descrizione
<code>hbase.oldwals.cleaner.thread.size</code>		Il numero di thread per pulire i WALs nella directory oldWALs.

Con Amazon EMR versione 5.17.0 e precedenti, l'operazione del processo di eliminazione può influenzare le prestazioni delle query durante l'esecuzione di carichi di lavoro gravosi, pertanto ti consigliamo di abilitare il processo di eliminazione solo durante gli orari non di punta. Il processo di pulitura dispone dei seguenti comandi shell HBase:

- `cleaner_chore_enabled` esegue la query se il processo di pulitura è abilitato.
- `cleaner_chore_run` esegue manualmente il processo di pulitura per rimuovere file.
- `cleaner_chore_switch` abilita o disabilita il processo di pulitura e restituisce lo stato precedente del processo di pulitura. Ad esempio, `cleaner_chore_switch true` abilita il processo di pulitura.

Proprietà per l'ottimizzazione delle prestazioni HBase su Amazon S3

I seguenti parametri possono essere regolati per ottimizzare le prestazioni del carico di lavoro durante l'utilizzo di HBase su Amazon S3.

Proprietà di configurazione	Valore predefinito	Descrizione
<code>hbase.bucketcache.size</code>	8,192	La quantità di spazio su disco, in MB, riservata su archivi istanza Amazon EC2 del server della regione e volumi EBS per l'archiviazione BucketCache. L'impostazione si applica a tutte le istanze del server della regione. Dimensioni BucketCache più grandi corrispondono generalmente a prestazioni migliori

Proprietà di configurazione	Valore predefinito	Descrizione
<code>hbase.hregion.memstore.flush.size</code>	134217728	Il limite di dati, in byte, in corrispondenza del quale viene attivato uno svuotamento memstore in Amazon S3.
<code>hbase.hregion.memstore.block.multiplier</code>	4	Un moltiplicatore che determina il limite superiore MemStore in corrispondenza del quale gli aggiornamenti vengono bloccati. Se il MemStore supera <code>hbase.hregion.memstore.flush.size</code> moltiplicato per questo valore, gli aggiornamenti vengono bloccati. Svuotamenti e compattazioni MemStore possono verificarsi per sbloccare gli aggiornamenti.
<code>hbase.hstore.blockingStoreFiles</code>	10	Il numero massimo di StoreFile contenuti in uno store prima che gli aggiornamenti vengono bloccati.
<code>hbase.hregion.max.filesize</code>	10737418240	Le dimensioni massime di una regione prima che venga divisa.

Chiusura e ripristino di un cluster senza perdita di dati

Per arrestare un cluster Amazon EMR senza perdere i dati che non sono stati scritti in Amazon S3 occorre svuotare la cache MemStore in Amazon S3 per scrivere nuovi file dello store. In primo luogo, dovrai disabilitare tutte le tabelle. La seguente configurazione di fase può essere utilizzata quando si

aggiunge una fase al cluster. Per ulteriori informazioni, consulta [Utilizzo di fasi mediante la AWS CLI e la console](#) nella Guida per la gestione di Amazon EMR.

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

In alternativa, puoi eseguire direttamente il seguente comando bash.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

Dopo aver disattivato tutte le tabelle, svuotare la tabella `hbase:meta` utilizzando la shell HBase e il seguente comando.

```
flush 'hbase:meta'
```

In seguito, puoi eseguire uno script della shell fornito sul cluster Amazon EMR per svuotare la cache di MemStore. Puoi aggiungerlo come una fase o eseguirlo direttamente utilizzando la AWS CLI sul cluster. Poiché lo script disabilita tutte le tabelle HBase, il MemStore in ciascun server regione viene svuotato in Amazon S3. Se lo script viene completato correttamente, i dati vengono mantenuti in Amazon S3 e il cluster può essere terminato.

Per riavviare un cluster con gli stessi dati HBase, specifica lo stesso percorso Amazon S3 del cluster precedente nella AWS Management Console o utilizzando la proprietà di configurazione `hbase.rootdir`.

Utilizzo della shell HBase

Dopo aver creato un cluster HBase, la fase successiva è connettersi ad HBase per poter iniziare a leggere e scrivere i dati (le scritture di dati non sono supportate su una replica di lettura al cluster). Puoi utilizzare la [shell HBase](#) per provare i comandi.

Per aprire la shell HBase

1. Utilizza SSH per connetterti al server master nel cluster HBase. Per informazioni su come connettersi al nodo master tramite SSH, consulta la sezione relativa alla [Connessione al nodo master tramite SSH](#) nella Guida per la gestione di Amazon EMR.
2. Esegui `hbase shell`. La shell HBase si apre con una richiesta simile a quella seguente.

```
hbase(main):001:0>
```

Puoi inviare comandi shell HBase dal prompt. Per ulteriori informazioni sui comandi shell e su come chiamarli, digita `help` al prompt HBase e premi Invio.

Creare una tabella

Il comando seguente consente di creare una tabella denominata 't1' che dispone di una singola famiglia di colonne denominata 'f1'.

```
hbase(main):001:0>create 't1', 'f1'
```

Inserimento di un valore

Il comando seguente consente di inserire il valore 'v1' per la riga 'r1' nella tabella 't1' e colonna 'f1'.

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

Ottenimento di un valore

Il comando seguente consente di ottenere i valori per la riga 'r1' nella tabella 't1'.

```
hbase(main):001:0>get 't1', 'r1'
```

Eliminazione di una tabella

Il comando seguente elimina ed elimina la tabella 't1'.

```
hbase(main):001:0>drop 'ns1:t1',false
```

Il valore booleano corrisponde a se si desidera archiviare o meno la tabella, quindi è possibile impostarla su `true` se vuoi salvarlo. Puoi anche eseguire `drop 'ns1:t1'` senza alcun valore booleano per archiviare la tabella.

Accesso a tabelle HBase con Hive

HBase e [Apache Hive](#) sono strettamente integrati. Ciò consente di eseguire carichi di lavoro MPP (Massively Parallel Processing) direttamente sui dati archiviati in HBase. Per utilizzare Hive con HBase, puoi generalmente avviarli sullo stesso cluster. Puoi, tuttavia, avviare Hive e HBase su cluster separati. L'esecuzione separata di HBase e Hive su cluster diversi può migliorare le prestazioni in quanto ogni applicazione può usare le risorse del cluster in modo più efficiente.

Le procedure seguenti mostrano come connettersi ad HBase su un cluster utilizzando Hive.

Note

Puoi connettere un cluster Hive solo a un singolo cluster HBase.

Per connettere Hive ad HBase

1. Creare cluster separati con Hive e HBase installati o creare un singolo cluster con HBase e Hive installati.
2. Se si stanno utilizzando cluster separati, modificare i gruppi di sicurezza in modo che le porte HBase e Hive siano aperte tra questi due nodi master.
3. Utilizzare SSH per connettersi al nodo master per il cluster con Hive installato. Per maggiori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
4. Avviare la shell Hive con il seguente comando.

```
hive
```

5. (Facoltativo) Questa operazione non è necessaria se HBase e Hive si trovano sullo stesso cluster. Collegare il client HBase sul cluster Hive al cluster HBase contenente i dati. Nell'esempio seguente, *public-DNS-name* viene sostituito dal nome DNS pubblico del nodo master del cluster HBase, ad esempio: `ec2-50-19-76-67.compute-1.amazonaws.com`.

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. Continuare a eseguire query Hive sui dati HBase come desiderato o passare alla procedura successiva.

Per accedere ai dati HBase da Hive

- Dopo che la connessione tra i cluster Hive e HBase è stata effettuata (come mostrato nella procedura precedente), puoi accedere ai dati archiviati nel cluster HBase creando una tabella esterna in Hive.

L'esempio seguente, quando eseguito dal prompt Hive sul nodo master, consente di creare una tabella esterna che fa riferimento a dati archiviati in una tabella HBase denominata `inputTable`. Puoi quindi fare riferimento a `inputTable` nelle istruzioni Hive per eseguire una query e modificare i dati memorizzati nel cluster 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 ;
```

Per un caso d'uso e un esempio più avanzati che combinano HBase e Hive, consulta l'articolo del Blog sui Big Data AWS [Combine NoSQL and massively parallel analytics using Apache HBase and Apache Hive on Amazon EMR \(Combinazione di NoSQL e analisi dei dati MPP \(Massively Parallel Processing\) utilizzando Apache HBase e Apache Hive su Amazon EMR\)](#).

Utilizzo di snapshot HBase

HBase utilizza una funzionalità [snapshot](#) integrata per creare backup leggeri di tabelle. Nel cluster EMR, questi backup possono essere esportati in Amazon S3 utilizzando EMRFS. Puoi creare uno snapshot sul nodo master usando la shell HBase. In questo argomento viene mostrato come eseguire questi comandi in maniera interattiva con la shell o tramite una fase utilizzando `command-runner.jar` con AWS CLI o AWS SDK for Java. Per ulteriori informazioni su altri tipi di backup HBase, consulta la sezione [HBase backup \(Backup HBase\)](#) nella documentazione di HBase.

Creazione di uno snapshot utilizzando una tabella

```
hbase snapshot create -n snapshotName -t tableName
```

Utilizzando `command-runner.jar` da AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \  
--steps Name="HBase Shell Step",Jar="command-runner.jar",\  
Args=[ "hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName"]
```

AWS SDK for Java

```
HadoopJarStepConfig hbaseSnapshotConf = new HadoopJarStepConfig()  
.withJar("command-runner.jar")  
.withArgs("hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName");
```

Note

Se il nome dello snapshot non è univoco, l'operazione di creazione non riesce e restituisce un codice di -1 o 255, ma non viene visualizzato alcun messaggio di errore che indica cosa si è verificato. Per usare lo stesso nome snapshot, eliminarlo e quindi ricrearlo.

Eliminazione di uno snapshot

```
hbase shell  
>> delete_snapshot 'snapshotName'
```

Visualizzazione delle informazioni relative allo snapshot

```
hbase snapshot info -snapshot snapshotName
```

Esportazione di uno snapshot in Amazon S3

Important

Se non si specifica un valore `-mappers` durante l'esportazione di uno snapshot, HBase utilizza un calcolo arbitrario per determinare il numero di mappatori. A seconda delle

dimensioni della tabella, questo valore può essere molto grande e influenzare negativamente i processi in esecuzione durante l'esportazione. Per questo motivo, ti consigliamo di specificare il parametro `-mappers`, il parametro `-bandwidth` (che specifica l'utilizzo della larghezza di banda in MB al secondo) o entrambi per limitare le risorse del cluster utilizzate dall'operazione di esportazione. In alternativa, puoi eseguire l'operazione di esportazione dello snapshot durante un periodo di basso utilizzo.

```
hbase snapshot export -snapshot snapshotName \
-copy-to s3://bucketName/folder -mappers 2
```

Utilizzando `command-runner.jar` da AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "export", "-snapshot", "snapshotName", "-copy-to", "s3://bucketName/folder", "-mappers", "2", "-bandwidth", "50"]
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
.withJar("command-runner.jar")
.withArgs("hbase", "snapshot", "export",
"-snapshot", "snapshotName", "-copy-to",
"s3://bucketName/folder",
"-mappers", "2", "-bandwidth", "50");
```

Importazione di uno snapshot da Amazon S3

Benché si tratti di un'importazione, l'opzione HBase utilizzata qui è ancora `export`.

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

Utilizzando `command-runner.jar` da 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");
```

Ripristino di una tabella da snapshot all'interno della shell HBase

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase non supporta attualmente tutti i comandi della snapshot trovati nella shell HBase. Ad esempio, non esiste un'opzione della riga di comando HBase per ripristinare una snapshot, pertanto è necessario ripristinarla all'interno di una shell. Questo significa che `command-runner.jar` deve eseguire un comando Bash.

Note

Poiché il comando utilizzato qui è `echo`, è possibile che il comando shell non riesca anche se il comando eseguito da Amazon EMR restituisce un codice di uscita 0. Se scegli di eseguire un comando della shell come una fase, controlla i log della fase.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Di seguito è riportata la fase utilizzando AWS CLI. Innanzitutto, crea il seguente file `snapshot.json`:

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$disable \"tableName\"; restore_snapshot \"snapshotName\"; enable \"tableName\" | hbase shell\"'],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./snapshot.json
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseRestoreSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash", "-c", "echo '$disable \"tableName\"; restore_snapshot \"snapshotName \"; enable \"snapshotName\" | hbase shell");
```

Configurazione di HBase

Anche se la maggior parte delle applicazioni funziona correttamente con le impostazioni predefinite, puoi scegliere di modificare le impostazioni di configurazione HBase. A tale scopo, utilizza le proprietà delle classificazioni di configurazione HBase. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

L'esempio seguente consente di creare un cluster con una directory principale HBase alternativa basata su un file di configurazione, `myConfig.json`, archiviato in Amazon S3.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --release-label emr-5.36.1 --applications Name=HBase \  
--instance-type m5.xlarge --instance-count 3 --configurations https://s3.amazonaws.com/  
mybucket/myfolder/myConfig.json
```

Il file `myConfig.json` specifica la proprietà `hbase.rootdir` per la classificazione della configurazione `hbase-site` come mostrato nel seguente esempio. Sostituisci `ip-XXX-XX-XX-XXX.ec2.internal` con il nome host DNS interno del nodo master del cluster.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/  
myCustomHBaseDir"  
    }  
  }  
]
```

Note

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Modifiche apportate all'allocazione di memoria in YARN

Poiché HBase non è in esecuzione come un'applicazione YARN, è necessario ricalcolare la quantità di memoria allocata per YARN e le relative applicazioni. Ciò risulta in una riduzione della memoria complessiva disponibile in YARN se HBase è installato. Durante la pianificazione occorre tenere conto di ciò per co-ubicare le applicazioni YARN e HBase sugli stessi cluster. I tipi di istanze con meno di 64 GB di memoria dispongono della metà della memoria disponibile in NodeManager, che viene quindi allocata ad HBase RegionServer. Per i tipi di istanze con memoria superiore a 64 GB, la memoria HBase RegionServer è limitata a 32 GB. Come regola generale, la memoria impostazione YARN è un multiplo della memoria dell'attività del riduttore MapReduce.

Le tabelle in [Valori predefiniti per le impostazioni di configurazione delle attività](#) mostrano le modifiche alle impostazioni YARN in base alla memoria necessaria per HBase.

Numeri di porta HBase

Alcuni numeri di porta scelti per HBase differiscono dall'impostazione predefinita. Di seguito sono riportate le interfacce e le porte per HBase su Amazon EMR.

Porte HBase

Interfaccia	Porta	Protocollo
HMaster	16000	TCP
HMaster UI	16010	HTTP
RegionServer	16020	TCP
RegionServer Info	16030	HTTP
Server REST	8070	HTTP
IU REST	8085	HTTP
Server Thrift	9090	TCP
Thrift server UI	9095	HTTP

Important

La `kms-http-port` è 9700 e la `kms-admin-port` è 9701 in Amazon EMR versione 4.6.0 e successive.

Impostazioni sito HBase da ottimizzare

Puoi impostare una qualsiasi o tutte le impostazioni sito HBase per ottimizzare il cluster HBase per il carico di lavoro dell'applicazione. Ti consigliamo le seguenti impostazioni come punto di partenza nella tua indagine.

zookeeper.session.timeout

Il timeout predefinito è 40 secondi (40000 ms). Se un server della regione si arresta in modo anomalo, questo è il tempo impiegato dal server master per notare l'assenza del server della regione e avviare il ripristino. Per accelerare il ripristino del server master, puoi ridurre questo valore a un periodo di tempo più breve. Nell'esempio seguente vengono utilizzati 30 secondi o 30000 ms:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

hbase.regionserver.handler.count

Definisce il numero di thread che vengono mantenuti aperti dal server della regione per servire richieste a tabelle. L'impostazione predefinita di 10 è bassa, in modo da impedire agli utenti di interrompere i propri server della regione durante l'utilizzo di buffer di scrittura di grandi dimensioni con un numero elevato di client simultanei. La regola generale è mantenere questo numero basso quando il payload per richiesta si avvicina all'intervallo di MB (put grandi, scan utilizzando una cache di grandi dimensioni) e alto quando il payload è di dimensioni ridotte (get, put piccoli, ICV, delete). Nell'esempio seguente il numero di thread aperti viene aumentato a 30:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

hbase.hregion.max.filesize

Questo parametro controlla le dimensioni, in byte, delle singole regioni. Per impostazione predefinita, è impostato su 1073741824. Nel caso di scrittura di una quantità elevata di dati nel cluster HBase che causa frequenti suddivisioni, puoi aumentare queste dimensioni per rendere più grandi le singole

regioni. Ciò consente di ridurre le suddivisioni ma richiede più tempo per bilanciare il carico delle regioni da un server a un altro.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.max.filesize": "1073741824"
    }
  }
]
```

hbase.hregion.memstore.flush.size

Questo parametro controlla le dimensioni massime di memstore, in byte, prima che venga svuotato nel disco. Per impostazione predefinita, tale valore è 134217728. Se il carico di lavoro è costituito da brevi picchi di operazioni di scrittura, potrebbe essere necessario aumentare questo limite in modo che tutte le scritture rimangano in memoria durante il picco e vengano svuotate nel disco più tardi. Ciò può incrementare le prestazioni durante i picchi.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

Visualizzazione dell'interfaccia utente di HBase

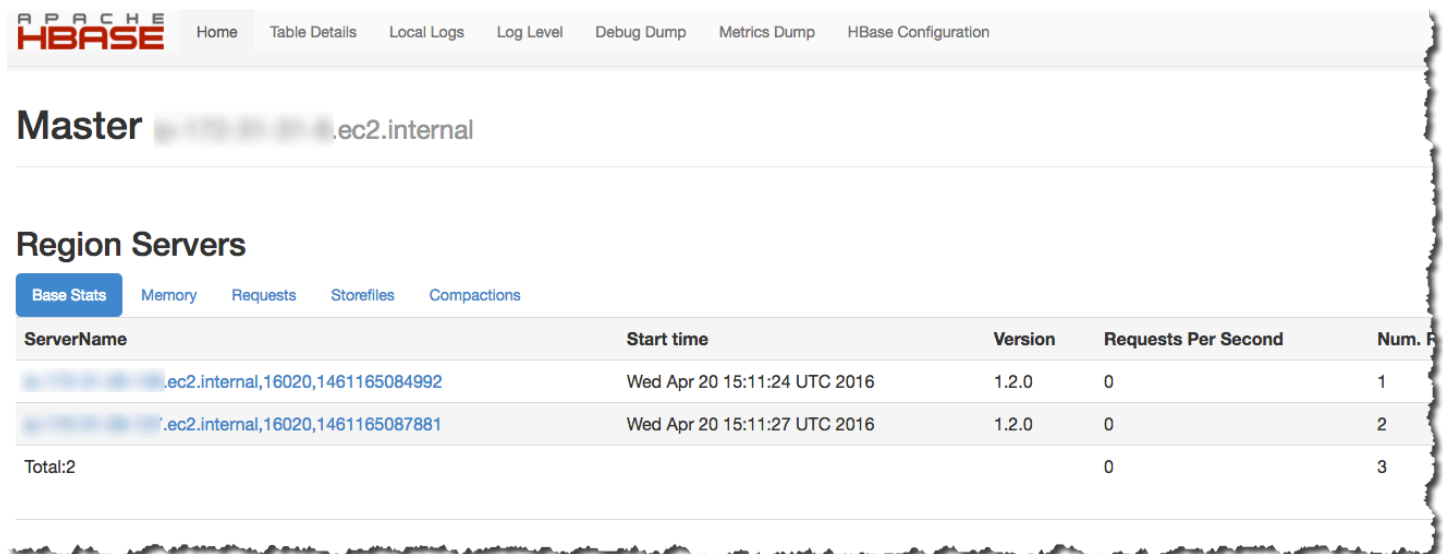
Note

L'interfaccia utente di HBase utilizza connessioni HTTP non sicure per impostazione predefinita. Per abilitare HTTP sicuro (HTTPS), imposta la proprietà `hbase.ssl.enabled` per la classificazione `hbase-site` su `true` nella tua [configurazione di HBase](#). Per ulteriori informazioni sull'utilizzo di HTTP sicuro (HTTPS) per l'interfaccia utente Web HBase, consulta la [Guida di riferimento di Apache HBase](#).

HBase offre un'interfaccia utente basata sul Web che puoi utilizzare per monitorare il cluster HBase. Quando esegui HBase su Amazon EMR, l'interfaccia Web viene eseguita sul nodo master e può essere visualizzata utilizzando l'inoltro porta, noto anche come creazione di un tunnel SSH.

Visualizzazione dell'interfaccia utente HBase

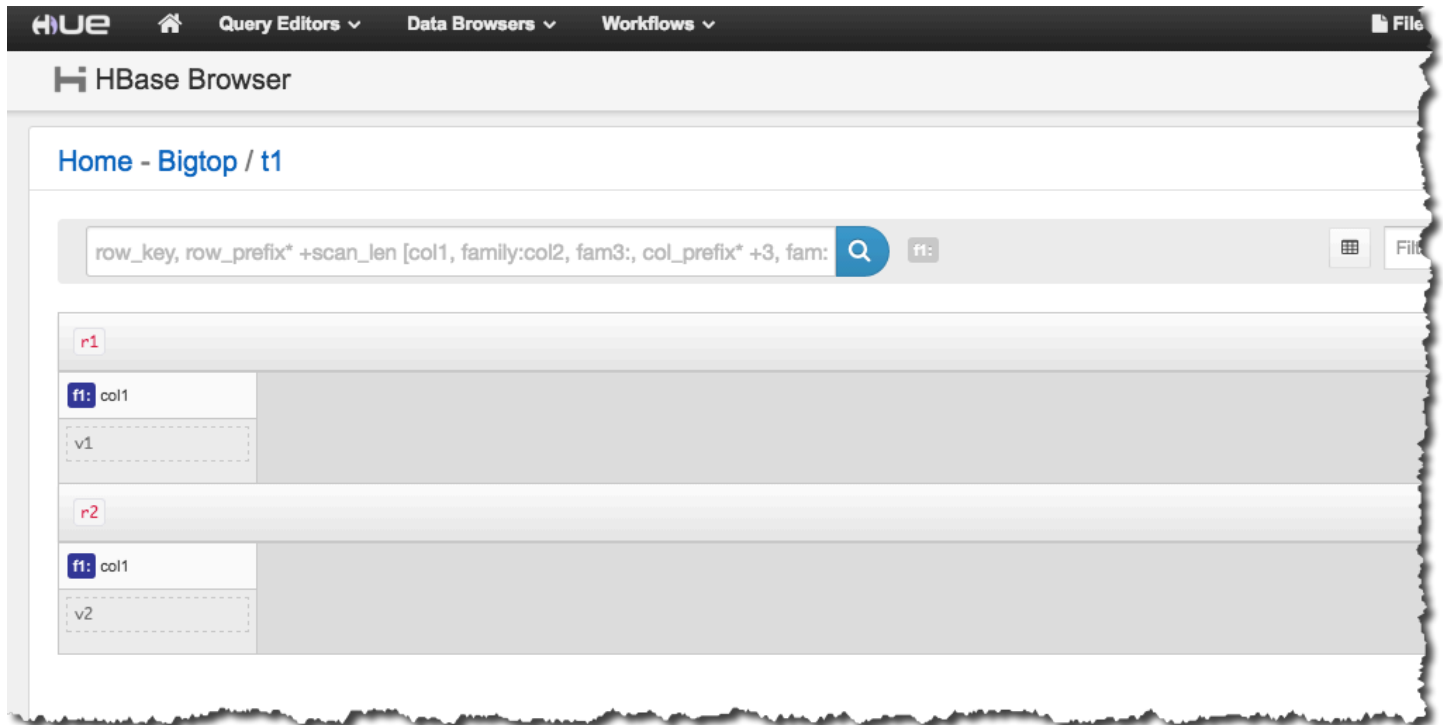
1. Utilizzare SSH per eseguire il tunneling nel nodo master e creare una connessione sicura. Per ulteriori informazioni, consulta la sezione relativa a [Opzione 2, Parte 1: Configurazione di un tunnel SSH sul nodo principale utilizzando l'inoltro porta dinamico](#) nella Guida alla gestione di Amazon EMR.
2. Installare un browser Web con uno strumento proxy, ad esempio il plug-in FoxyProxy per Firefox, per creare un proxy SOCKS per domini AWS. Per ulteriori informazioni, consulta la sezione relativa a [Opzione 2, Parte 2: Configurazione delle impostazioni proxy per visualizzare siti Web ospitati sul nodo master](#) nella Guida alla gestione di Amazon EMR.
3. Con il proxy impostato e la connessione SSH aperta, puoi visualizzare l'interfaccia utente HBase aprendo una finestra del browser con `http://master-public-dns-name:16010/master-status`, dove *master-public-dns-name* è l'indirizzo DNS pubblico del nodo master del cluster.



The screenshot shows the Apache HBase web interface. At the top, there is a navigation bar with links: Home, Table Details, Local Logs, Log Level, Debug Dump, Metrics Dump, and HBase Configuration. Below the navigation bar, the page title is "Master" followed by a blurred IP address and ".ec2.internal". Underneath, there is a section titled "Region Servers" with a sub-section "Base Stats" selected. Below this, there is a table with the following data:

ServerName	Start time	Version	Requests Per Second	Num. P
<i>blurred</i> .ec2.internal,16020,1461165084992	Wed Apr 20 15:11:24 UTC 2016	1.2.0	0	1
<i>blurred</i> .ec2.internal,16020,1461165087881	Wed Apr 20 15:11:27 UTC 2016	1.2.0	0	2
Total:2			0	3

Puoi anche visualizzare HBase in Hue. Ad esempio, di seguito viene visualizzata la tabella, t1, creata in [Utilizzo della shell HBase](#):



Per ulteriori informazioni su Hue, consulta [Hue](#).

Visualizzazione di file di log HBase

Come parte del suo funzionamento, HBase scrive file di log con dettagli sulle impostazioni di configurazione, azioni daemon ed eccezioni. Questi file di log possono essere utili per il debug dei problemi riscontrati con HBase, nonché per il monitoraggio delle prestazioni.

Se configuri il cluster per mantenere i file di log in Amazon S3 devi tenere a mente che i log vengono scritti in Amazon S3 ogni cinque minuti, pertanto potrebbe verificarsi un lieve ritardo prima che i file di log più recenti siano disponibili.

Per visualizzare i file di log di HBase per il nodo master

- È possibile visualizzare i log di HBase correnti utilizzando SSH per la connessione al nodo master e accedendo alla directory `/var/log/hbase`. Questi file di log non sono disponibili dopo che il cluster viene terminato a meno che non si abiliti la registrazione in Amazon S3 quando il cluster viene avviato.

Visualizzazione dei log di HBase su Amazon S3

- Per accedere ai file di log di HBase e ad altri file di log del cluster su Amazon S3 e fare in modo che siano disponibili dopo che il cluster viene terminato, specifica un bucket Amazon S3 per ricevere questi file di log quando crei il cluster. Questa operazione viene eseguita utilizzando l'opzione `--log-uri`. Per ulteriori informazioni su come abilitare la registrazione per il cluster, consulta la sezione sulla [Configurazione della registrazione e del debugging \(facoltativo\)](#) nella Guida alla gestione di Amazon EMR.

Monitoraggio di HBase con Ganglia

Il progetto open source Ganglia è un sistema scalabile e distribuito, progettato per monitorare i cluster e le griglie riducendo al minimo l'impatto sulle loro prestazioni. Quando abiliti Ganglia sul cluster, puoi generare report e visualizzare le prestazioni del cluster nel suo insieme, nonché esaminare le prestazioni di singole istanze nodi. Per ulteriori informazioni sul progetto open source Ganglia, consulta <http://ganglia.info/>. Per ulteriori informazioni sull'uso di Ganglia con cluster Amazon EMR, consulta [Ganglia](#).

Dopo che il cluster viene avviato con Ganglia configurato, puoi accedere ai grafici e ai report Ganglia utilizzando l'interfaccia grafica in esecuzione sul nodo master.

Ganglia archivia i file di log nella directory `/mnt/var/lib/ganglia/rrds/` del nodo master. In versioni precedenti di Amazon EMR è possibile che i file di log vengano archiviati nella directory `/var/log/ganglia/rrds/`.

Per configurare un cluster per Ganglia e HBase utilizzando AWS CLI

- Utilizzare un comando `create-cluster` simile al seguente:

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \  
--applications Name=HBase Name=Ganglia --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3
```

Note

Se il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza Amazon EC2 non esistono, si verifica un errore. Utilizza il comando `aws emr create-default-roles` per crearli e riprovare.

Per ulteriori informazioni, consulta [Comandi Amazon EMR nella AWS CLI](#).

Per visualizzare parametri HBase nell'interfaccia Web Ganglia

1. Utilizzare SSH per eseguire il tunneling nel nodo master e creare una connessione sicura. Per ulteriori informazioni, consulta la sezione relativa a [Opzione 2, Parte 1: Configurazione di un tunnel SSH sul nodo master utilizzando l'inoltro porta dinamico](#) nella Guida alla gestione di Amazon EMR.
2. Installare un browser Web con uno strumento proxy, ad esempio il plug-in FoxyProxy per Firefox, per creare un proxy SOCKS per domini AWS. Per ulteriori informazioni, consulta la sezione relativa a [Opzione 2, Parte 2: Configurazione delle impostazioni proxy per visualizzare siti Web ospitati sul nodo master](#) nella Guida alla gestione di Amazon EMR.
3. Con il proxy impostato e la connessione SSH aperta, puoi visualizzare i parametri Ganglia aprendo una finestra del browser con `http://master-public-dns-name/ganglia/`, dove *master-public-dns-name* è l'indirizzo DNS pubblico del server master nel cluster HBase.

Per visualizzare i file di log Ganglia sul nodo master

- Se il cluster è ancora in esecuzione, puoi accedere ai file di log utilizzando SSH per connetterti al nodo master e passare alla directory `/mnt/var/lib/ganglia/rdds/`. Su EMR 3.x, passa alla directory `/var/log/ganglia/rdds`. Per maggiori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.

Visualizzazione dei file di log Ganglia su Amazon S3

- I file di log Ganglia non vengono scritti automaticamente in Amazon S3 anche quando la registrazione per il cluster è abilitata. Per visualizzare i file di log Ganglia in Amazon S3, è

necessario eseguire manualmente il push dei log da `/mnt/var/lib/ganglia/rrds/` al bucket S3.

Migrazione da versioni HBase precedenti

Per eseguire la migrazione dei dati da una versione HBase precedente, consulta [Upgrading \(Aggiornamento\)](#) e [HBase version number and compatibility \(Numero di versione HBase e compatibilità\)](#) nella Apache HBase Reference Guide. Potrebbe essere necessario prestare particolare attenzione ai requisiti per l'aggiornamento da versioni di HBase precedenti a 1.0.

Cronologia dei rilasci di HBase

La tabella seguente indica la versione di HBase inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di HBase

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-ope

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
		rator-tools, zookeeper-client, zookeeper-server
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-ftpfs-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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-6.11.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-6.10.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-6.10.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.36.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.6.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.33.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.28.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.28.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.26.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.25.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.24.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.14.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.13.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.0	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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.11.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.11.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.10.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.8.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.3.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-5.2.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.1.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, 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-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
emr-4.8.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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-4.8.0	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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
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
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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HBase	Componenti installati con HBase
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-httpfs-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

HCatalog Apache

HCatalog è uno strumento che ti consente di accedere alle tabelle del metastore Hive all'interno di Pig, Spark SQL e/o delle applicazioni MapReduce personalizzate. HCatalog dispone di un'interfaccia REST e di un client a riga di comando che ti consentono di creare tabelle o effettuare altre operazioni. Quindi, puoi avviare la scrittura delle applicazioni per accedere alle tabelle tramite le librerie HCatalog. Per ulteriori informazioni, consulta la sezione relativa all'[Uso di HCatalog](#). HCatalog è incluso nel rilascio di Amazon EMR versione 4.4.0 e successive.

HCatalog su Amazon EMR versione 5.8.0 e successive supporta l'utilizzo di AWS Glue Data Catalog come metastore per Hive. Per ulteriori informazioni, consulta la sezione relativa all'[Utilizzo di AWS Glue Data Catalog come metastore per Hive](#).

La seguente tabella indica la versione di HCatalog inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con HCatalog.

Per la versione dei componenti installati con HCatalog in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di HCatalog per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-6.14.0	HCatalog 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

La seguente tabella indica la versione di HCatalog inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con HCatalog.

Per la versione dei componenti installati con HCatalog in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di HCatalog per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-5.36.1	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

Argomenti

- [Creazione di un cluster con HCatalog](#)
- [Uso di HCatalog](#)
- [Esempio: Creare una tabella HCatalog e scrivervi con Pig](#)
- [Cronologia dei rilasci di HCatalog](#)

Creazione di un cluster con HCatalog

Anche se HCatalog è incluso nel progetto Hive, devi installarlo come applicazione separata.

Per avviare un cluster con lo strumento HCatalog installato tramite la console

La procedura seguente consente di creare un cluster con lo strumento HCatalog installato. Per ulteriori informazioni sulla creazione di cluster tramite la console, tra cui le Advanced Options (Opzioni avanzate), consulta la sezione relativa alla [Pianificazione e configurazione dei cluster](#) della Guida per la gestione di Amazon EMR.

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegliere Create cluster (Crea cluster) per utilizzare l'opzione Quick Create (Creazione rapida).
3. Per il campo Software Configuration (Configurazione software), scegliere Amazon Release Version emr-4.4.0 (Amazon versione emr-4.4.0) o versioni successive.
4. Nel campo Select Applications (Seleziona applicazioni), scegliere All Applications (Tutte le applicazioni) o HCatalog.
5. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Per avviare un cluster con HCatalog tramite AWS CLI

- Creare il cluster con il seguente comando:

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
aws emr create-cluster --name "Cluster with Hcat" --release-label emr-5.36.1 \  
--applications Name=HCatalog --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Uso di HCatalog

Puoi utilizzare HCatalog all'interno di diverse applicazioni che utilizzano il metastore Hive. Negli esempi di questa sezione viene mostrato come creare una tabella e utilizzarla nel contesto di Pig e Spark SQL.

Disabilitazione della scrittura diretta durante l'utilizzo di HCatalog HStorer

Ogni volta che un'applicazione utilizza [HCatStorer](#) per scrivere su una tabella HCatalog archiviata in Amazon S3, disabilita la caratteristica di scrittura diretta di Amazon EMR. Ad esempio, disabilita la scrittura diretta quando utilizzi il comando STORE Pig o durante l'esecuzione di processi Sqoop che scrivono le tabelle HCatalog su Amazon S3. Puoi disabilitare la caratteristica di scrittura diretta impostando le configurazioni `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` su `false`. Nell'esempio seguente viene illustrato come impostare queste configurazioni tramite Java.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

Creazione di una tabella tramite la CLI HCat e uso dei dati in Pig

Crea lo script seguente, `impressions.q`, sul 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');
```

Esegui lo script tramite la CLI HCat:

```
% 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
```

Apri la shell Grunt e accedi ai dati in `impressions`:

```
% pig -useHCatalog -e "A = LOAD 'impressions' USING
  org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 5;
dump B;"
<snip>
(1239610346000,m9nwdo67Nx6q2kI25qt50n7peICfUM,omkxkaRpNhGPDucAiBErSh1cs0MThC,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 6.0; FunWebProducts; GTB6; SLCC1; .NET CLR
  2.0.50727; Media Center PC
  5.0; .NET,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,Irpv3oiu0I5QNQiWSSIIshrLdo9cM1,i1LDq44LRSJF0hbmhB8Gk7k9gMWtBq,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239611007000,q2Awfnpe0JAvhInaIp0VGx9KTS0oPO,s3HvTf1PB8JIE0IuM6h0EebWwP0tJV,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239610398000,c362vpAB0soPKGHRs43cj6TRwNe0Gn,jeas5nXbQInGAgFB8jlkhnprN6cMw7,cartoonnetwork.com
  (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB6; .NET CLR
  1.1.4322),k96n5PnUmwHKfiUI0TFP0TNMFADgh9,51.131.29.87,2009-04-13-08-05)
7120 [main] INFO  org.apache.pig.Main - Pig script completed in 7 seconds and 199
  milliseconds (7199 ms)
16/03/08 23:17:10 INFO pig.Main: Pig script completed in 7 seconds and 199 milliseconds
  (7199 ms)
```

Accesso alla tabella tramite Spark SQL

In questo esempio viene creato un DataFrame Spark dalla tabella creata nel primo esempio e vengono mostrate le prime 20 righe:

```
% spark-shell --jars /usr/lib/hive-hcatalog/share/hcatalog/hive-hcatalog-core-1.0.0-
  amzn-3.jar
<snip>
scala> val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);
scala> val df = hiveContext.sql("SELECT * FROM impressions")
scala> df.show()
<snip>
16/03/09 17:18:46 INFO DAGScheduler: ResultStage 0 (show at <console>:32) finished in
  10.702 s
16/03/09 17:18:46 INFO DAGScheduler: Job 0 finished: show at <console>:32, took
  10.839905 s
```

```

+-----+-----+-----+-----+
+-----+-----+-----+-----+
|requestbegintime|          adid|          impressionid|          referrer|
|  useragent|          usercookie|          ip|          dt|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
|  1239610346000|m9nwd067Nx6q2kI25...|omkxkaRpNhGPDucAi...|cartoonnetwork.com|
Mozilla/4.0 (comp...|wcVWWTascoPbGt6bd...|69.191.224.234|2009-04-13-08-05|
|  1239611000000|NjriQjd0DgWBKnkGJ...|AWtXPKfawG0aNeL90...|cartoonnetwork.com|
Mozilla/4.0 (comp...|0aMU1F2gE4CtADVHA...| 57.34.133.110|2009-04-13-08-05|
|  1239610462000|Irpv3oiu0I5QNQiW...|i1LDq44LRSJF0hbmh...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239611007000|q2Awfnpe0JAvhInaI...|s3HvTf1PB8JIE0IuM...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239610398000|c362vpAB0soPKGHR...|jeas5nXbQInGAgFB8...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610600000|cjbTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVL1625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610872000|h5bccHX6wJReDi1jL...|EFAWIiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdTXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
|  1239610348000|X8gISpUTSqh1A5reS...|TrFblGT99AgE75vuj...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610743000|kbKreLWB6QVueFrDm...|kVnxx9Ie2i30LTxFj...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610812000|9lX0SRpEi3bmEeTCu...|1B2sff99AEIwSuLVV...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610876000|lijjmCf2kuxfBTnjL...|AjvufgUtakUFcsIM9...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610941000|t8t8trgjNRPIlmxuD...|agu2u2TCdqWP08rAA...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610490000|OGRLPVNGxiGgrCmWL...|mJg2raBUpPrC80lUm...|          corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
|  1239610556000|OnJID12x0RXKPUgrD...|P7Pm2mPdW6w08KA3R...|          corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
|  1239610373000|WflsvKIg0qfIE5KwR...|TJHd1VBspNcua0XPn...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|  1239610768000|4MJR0XxiVCU1ueXKV...|10hGwmbvKf8ajoU8a...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|  1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|

```

```
| 1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
only showing top 20 rows

scala>
```

Esempio: Creare una tabella HCatalog e scrivervi con Pig

Puoi creare una tabella HCatalog e utilizzare Apache Pig per scrivervi sopra tramite HCatStorer utilizzando un'origine dati in Amazon S3. Per utilizzare HCatalog, occorre disabilitare la scrittura diretta, altrimenti l'operazione non verrà completata e non verrà inviato alcun errore. Imposta entrambe le configurazioni `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` su `false` tramite la classificazione `mapred-site` oppure manualmente dalla shell Grunt. Nell'esempio seguente viene mostrata una tabella creata tramite la CLI HCat, seguita dai comandi eseguiti nella shell Grunt per compilare la tabella a partire da un file di dati di esempio in Amazon S3.

Per eseguire questo esempio, [connettiti al nodo master tramite SSH](#).

Crea un file di script HCatalog, `wikicount.q`, con i contenuti seguenti, che creerà una tabella HCatalog denominata `wikicount`.

```
CREATE EXTERNAL TABLE IF NOT EXISTS wikicount(
  col1 string,
  col2 bigint
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001'
STORED AS ORC
LOCATION 's3://MyBucket/hcat/wikicount';
```

Utilizza un comando della CLI HCat per eseguire lo script dal file.

```
hcat -f wikicount.q
```

Quindi, avvia la shell Grunt con l'opzione `-useHCatalog`, imposta le configurazioni per disabilitare la scrittura diretta, carica i dati da un percorso S3 e scrivi i risultati sulla tabella `wikicount`.

```

pig -useHCatalog
SET mapred.output.direct.NativeS3FileSystem false;
SET mapred.output.direct.EmrFileSystem false;
A = LOAD 's3://support.elasticmapreduce/training/datasets/wikistats_tiny/' USING
  PigStorage(' ') AS (Site:chararray, page:chararray, views:int, total_bytes:long);
B = GROUP A BY Site;
C = FOREACH B GENERATE group as col1, COUNT(A) as col2;
STORE C INTO 'wikicount' USING org.apache.hive.hcatalog.pig.HCatStorer();

```

Cronologia dei rilasci di HCatalog

Nella seguente tabella sono elencati la versione di HCatalog inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di HCatalog

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-librar

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
		y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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-resource-manager, 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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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-resourcemanager, 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-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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-resource-manager, 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-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.2	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

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con 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-resourcemanager, 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-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di HCatalog	Componenti installati con HCatalog
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Apache Hive

Hive è un pacchetto open source di data warehouse e analisi eseguibile in un cluster Hadoop. Gli script Hive utilizzano un linguaggio di tipo SQL denominato Hive QL (linguaggio di query) che garantisce l'astrazione dei modelli di programmazione e supporta interazioni classiche di data warehouse. Hive ti consente di evitare le complessità inerenti alla scrittura di processi Tez basati su grafi aciclici diretti (DAG) o programmi MapReduce in un linguaggio di programmazione di livello inferiore come Java.

Hive estende il paradigma SQL includendo formati di serializzazione. Puoi inoltre personalizzare l'elaborazione di query mediante la creazione di uno schema di tabella corrispondente ai tuoi dati, senza la necessità di alterarli. A differenza di SQL (che supporta solo tipi di valore primitivi come date, numeri e stringhe), i valori nelle tabelle Hive sono elementi strutturati, ad esempio oggetti JSON, qualsiasi tipo di dati definiti dall'utente o qualsiasi funzione scritta in Java.

Per ulteriori informazioni su Hive, visita la pagina Web all'indirizzo <http://hive.apache.org/>.

La seguente tabella indica la versione di Hive inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Hive.

Per la versione dei componenti installati con Hive in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hive per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.14.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
		hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

La seguente tabella indica la versione di Hive inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Hive.

Per la versione dei componenti installati con Hive in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hive per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.36.1	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-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

A partire da Amazon EMR rilascio 5.18.0, è possibile utilizzare i repository degli artefatti di Amazon EMR per creare il codice di processo secondo le versioni esatte di librerie e dipendenze disponibili

con rilasci Amazon EMR specifici. Per ulteriori informazioni, consulta [Verifica delle dipendenze mediante il repository degli artifact di Amazon EMR](#).

Argomenti

- [Differenze e considerazioni per Hive su Amazon EMR](#)
- [Configurazione di un metastore esterno per Hive](#)
- [Utilizzo del driver JDBC Hive](#)
- [Miglioramento delle prestazioni Hive](#)
- [Utilizzo di Hive Live Long and Process \(LLAP\)](#)
- [Crittografia in Hive](#)
- [Cronologia dei rilasci di Hive](#)

Differenze e considerazioni per Hive su Amazon EMR

Differenze tra Apache Hive su Amazon EMR e Apache Hive

Questa sezione descrive le differenze tra Hive su Amazon EMR e le versioni predefinite di Hive disponibili su <http://svn.apache.org/viewvc/hive/branches/>.

Autorizzazione Hive

Amazon EMR supporta le [autorizzazioni Hive](#) per HDFS ma non per EMRFS e Amazon S3. I cluster Amazon EMR vengono eseguiti con l'autorizzazione disabilitata per impostazione predefinita.

Comportamento di unione dei file Hive con Amazon S3

Apache Hive unisce i piccoli file al termine di un processo di sola mappatura se `hive.merge.mapfiles` è impostato su `true`; l'unione viene attivata solo se le dimensioni medie dell'output del processo sono inferiori all'impostazione `hive.merge.smallfiles.avgsize`. Il comportamento di Amazon EMR Hive è esattamente lo stesso se il percorso di output finale si trova in HDFS. Se il percorso di output si trova in Amazon S3, il parametro `hive.merge.smallfiles.avgsize` viene ignorato. In tal caso, l'attività di unione viene sempre attivata se `hive.merge.mapfiles` è impostato su `true`.

Transazioni ACID e Amazon S3

Amazon EMR 6.1.0 e versioni successive supporta le transazioni Hive ACID (Atomicity, Consistency, Isolation, and Durability) in modo da conformarsi alle proprietà ACID di un database. Con questa funzione, puoi eseguire operazioni INSERT (INSERISCI), UPDATE (AGGIORNA), DELETE (ELIMINA) e MERGE (UNISCI) nelle tabelle gestite da Hive con dati in Amazon Simple Storage Service (Amazon S3).

Hive LLAP (Live Long e Process)

La [funzionalità LLAP](#), aggiunta nella versione 2.0 di Apache Hive predefinito, non è supportata in Hive 2.1.0 su Amazon EMR rilascio 5.0.

Amazon EMR versione 6.0.0 e versioni successive supporta la funzionalità Live Long and Process (LLAP) per Hive. Per ulteriori informazioni, consulta la sezione relativa all'[utilizzo di LLAP Hive](#).

Differenze in Hive tra le versioni Amazon EMR 4.x e 5.x

In questa sezione vengono descritte le differenze da considerare prima di migrare un'implementazione da Hive versione 1.0.0 su Amazon EMR versione 4.x a Hive 2.x su Amazon EMR versione 5.x.

Differenze e considerazioni operative

- Supporto aggiunto per [transazioni ACID \(Atomicity, Consistency, Isolation, and Durability\)](#): questa differenza tra Hive 1.0.0 su Amazon EMR 4.x e Apache Hive predefinito è stata eliminata.
- Scritture dirette su Amazon S3 eliminate: questa differenza tra Hive 1.0.0 su Amazon EMR e Apache Hive predefinito è stata eliminata. Hive 2.1.0 su Amazon EMR versione 5.x ora esegue operazioni di creazione, lettura e scrittura sui file temporanei archiviati in Amazon S3. Di conseguenza, per leggere e scrivere sulla stessa tabella, non è più necessario avviare creando una tabella temporanea nel file system HDFS locale del cluster. Con la funzione Versioni multiple dei bucket attivata, è necessario assicurarsi di gestire questi file temporanei come descritto di seguito.
- Gestione dei file temporanei con la funzione Versioni multiple dei bucket Amazon S3 attivata: quando vengono eseguite query Hive in cui la destinazione dei dati generati è Amazon S3, vengono creati molti file e directory temporanei. Si tratta di un nuovo comportamento, come descritto in precedenza. Se utilizzi i bucket S3 con funzione Versioni multiple, questi file temporanei creano disordine in Amazon S3 e, se non eliminati, generano costi. Modifica le regole del ciclo di

vita in modo che i dati con prefisso `/_tmp` vengano eliminati dopo poco tempo, ad esempio cinque giorni. Per ulteriori informazioni, consulta [Specifica di una configurazione del ciclo di vita](#).

- Log4j aggiornato a log4j 2: se utilizzi log4j, potresti dover modificare la configurazione di log per via di questo aggiornamento. Per dettagli, consulta [Apache log4j 2](#).

Differenze e considerazioni prestazionali

- Differenze di prestazioni con Tez: con Amazon EMR versione 5.x, il motore di esecuzione predefinito per Hive è Tez, anziché MapReduce. Tez fornisce prestazioni migliori per la maggior parte dei flussi di lavoro.
- Tabelle con molte partizioni: le query che generano un numero elevato di partizioni dinamiche possono non riuscire, mentre quelle che operano selezioni da tabelle con molte partizioni possono richiedere più tempo del previsto per l'esecuzione. Ad esempio, una selezione da 100.000 partizioni può richiedere 10 minuti o più.

Caratteristiche aggiuntive di Hive su Amazon EMR

Amazon EMR amplia Hive con nuove caratteristiche che supportano l'integrazione di Hive con altri servizi AWS, ad esempio la possibilità di leggere e scrivere su Amazon Simple Storage Service (Amazon S3) e DynamoDB.

Variabili in Hive

Puoi includere variabili nel tuo script utilizzando il segno del dollaro e le parentesi graffe.

```
add jar ${LIB}/jsonserde.jar
```

Puoi trasferire ad Hive i valori di queste variabili nella riga di comando tramite il parametro `-d`, come nell'esempio seguente:

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

Puoi anche trasferire i valori nelle fasi che eseguono gli script Hive.

Per trasferire i valori delle variabili nelle fasi Hive utilizzando la console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegli Crea cluster.
3. Nella sezione Steps (Fasi), in Add Step (Aggiungi fase), scegliere Hive Program (Programma Hive) dall'elenco e quindi Configure and add (Configura e aggiungi).
4. Nella finestra di dialogo Add Step (Aggiungi fase), specificare i parametri utilizzando la tabella seguente come guida e quindi scegliere Add (Aggiungi).

Campo	Azione
Percorso script S3*	Specifica l'URI in cui risiede lo script in Amazon S3. Il valore deve essere nel formato <i>NomeBucket /percorso/NomeScript</i> . Ad esempio: <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> .
Percorso di input S3	Facoltativamente, è possibile specificare l'URI in cui risiedono file di input in Amazon S3. Il valore deve essere nel formato <i>NomeBucket /percorso/</i> . Se specificato, verrà trasferito nello script Hive come parametro, denominato INPUT. Ad esempio: <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> .
Percorso di output S3	Facoltativamente, è possibile specificare l'URI in cui archiviare l'output in Amazon S3. Il valore deve essere nel formato <i>NomeBucket /percorso/</i> . Se specificato, verrà trasferito nello script Hive come parametro, denominato OUTPUT. Ad esempio: <code>s3://mybucket/hive-ads/output/</code> .
Argomenti	Facoltativamente, è possibile inserire un elenco di argomenti (stringhe separate da spazi) per il trasferimento a Hive. Se è stato definito un percorso variabile nello script Hive denominato <code>\${SAMPLE}</code> , ad esempio: <pre>CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde' WITH SERDEPROPERTIES ('paths'='requestBeginTime, requestEndTime, hostname') LOCATION '\${SAMPLE}/tables/impressions';</pre>

Campo	Azione
	<p>Per trasferire un valore per la variabile, digitare quanto segue nella finestra Arguments (Argomenti):</p> <pre>-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/ .</pre>
Operazione in caso di errore	<p>Determina le operazioni effettuate dal cluster in risposta a eventuali errori. I valori possibili sono per questa impostazione sono:</p> <ul style="list-style-type: none"> • Terminate cluster (Chiudi cluster): se la fase ha esito negativo, il cluster viene chiuso. Se vi sono abilitati sia la protezione dalla chiusura che il keep-alive, il cluster non verrà chiuso. • Cancel and wait (Annulla e attendi): se la fase ha esito negativo, quelle rimanenti vengono annullate. Se il keep-alive è abilitato, il cluster non verrà chiuso. • Continue (Continua): se la fase ha esito negativo, passa a quella successiva.

5. Selezionare i valori in base alle esigenze, quindi scegliere Create cluster (Crea cluster).

Per trasferire i valori delle variabili nelle fasi Hive utilizzando l'AWS CLI

Per trasferire i valori delle variabili nelle fasi di Hive tramite AWS CLI, utilizzare il parametro `--steps` e includere un elenco di argomenti.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \
--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/]
```

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Per trasferire i valori delle variabili nelle fasi Hive utilizzando l'SDK Java

- L'esempio seguente spiega come trasferire le variabili nelle fasi utilizzando l'SDK. Per ulteriori informazioni, consulta la sezione relativa alla [Classe StepFactory](#) nella Guida di riferimento alle API di AWS SDK for Java.

```
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")));
```

Query di Amazon EMR Hive per soddisfare schemi DynamoDB parziali

Amazon EMR Hive offre grande flessibilità nell'esecuzione di query in tabelle DynamoDB poiché consente di specificare un sottoinsieme di colonne in cui è possibile filtrare i dati, piuttosto che richiedere alla query di includere tutte le colonne. La tecnica di query a schema parziale è efficace quando, con schemi del database sparsi, vuoi filtrare i record in base a poche colonne, ad esempio applicando il filtro in base al timestamp.

L'esempio seguente mostra come utilizzare una query Hive per:

- Creazione di una tabella DynamoDB
- Seleziona un sottoinsieme di elementi (righe) in DynamoDB e restringere ulteriormente i dati in determinate colonne.
- Copia i dati risultati 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";

```

La tabella seguente mostra la sintassi di query per la selezione di qualsiasi combinazione di elementi da DynamoDB.

Esempio di query	Descrizione del risultato
SELECT * FROM <i>nome_tabella</i> ;	Seleziona tutti gli elementi (righe) da una determinata tabella e include i dati provenienti da tutte le colonne disponibili per tali elementi.
SELECT * FROM <i>nome_tabella</i> WHERE <i>nome_campo</i> = <i>valore</i> ;	Seleziona alcuni elementi (righe) da una determinata tabella e include i dati provenienti da tutte le colonne disponibili per tali elementi.
SELECT <i>nome_colonna1</i> , <i>nome_colonna2</i> , <i>nome_colonna3</i> FROM <i>nome_tabella</i> ;	Seleziona tutti gli elementi (righe) da una determinata tabella e include i dati provenienti da alcune colonne disponibili per tali elementi.
SELECT <i>nome_colonna1</i> , <i>nome_colonna2</i> , <i>nome_colonna3</i> FROM <i>nome_tabella</i> WHERE <i>nome_campo</i> = <i>valore</i> ;	Seleziona alcuni elementi (righe) da una determinata tabella e include i dati provenienti da alcune colonne disponibili per tali elementi.

Copia dei dati fra tabelle DynamoDB in diverse Regioni AWS

Amazon EMR Hive fornisce una proprietà `dynamodb.region` che è possibile impostare per ogni tabella DynamoDB. Quando `dynamodb.region` è impostato in modo diverso su due tabelle, la copia dei dati da una tabella all'altra avviene tra le regioni specificate.

L'esempio seguente mostra come creare una tabella DynamoDB con uno script Hive che imposta la proprietà `dynamodb.region`:

Note

Le proprietà delle regioni specifiche per tabella sostituiscono le proprietà globali di Hive.

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.region" = "eu-west-1",
    "dynamodb.throughput.read.percent" = ".1000",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

Impostazione dei valori di velocità effettiva DynamoDB per tabella

Amazon EMR Hive consente di configurare le impostazioni `writeThroughputPercent` e `readThroughputPercent` di DynamoDB per ogni tabella nella relativa definizione. Il seguente script Amazon EMR Hive mostra come impostare i valori di velocità effettiva. Per ulteriori informazioni sui valori di velocità effettiva di DynamoDB, consulta [Specifica dei requisiti di lettura e scrittura per le tabelle](#).

```
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");
```

Configurazione di un metastore esterno per Hive

Per impostazione predefinita, Hive registra le informazioni del metastore in un database MySQL sul file system del nodo primario. Il metastore contiene una descrizione della tabella e dei dati sottostanti su cui è costruita, compresi i nomi delle partizioni, i tipi di dati e così via. Quando un cluster termina, tutti i nodi del cluster vengono chiusi, incluso il nodo primario. Quando ciò accade, i dati locali vengono persi perché i file system dei nodi utilizzano una memorizzazione effimera. Se occorre che il metastore persista, è necessario creare un metastore esterno esistente all'esterno del cluster.

Sono disponibili due opzioni per il metastore esterno:

- AWS Glue Data Catalog (solo Amazon EMR rilascio 5.8.0 o successivi).

Per ulteriori informazioni, consulta [Utilizzo di AWS Glue Data Catalog come metastore Hive](#).

- Amazon RDS o Amazon Aurora.

Per ulteriori informazioni, consulta [Utilizzo di un database MySQL o Amazon Aurora esterno](#).

Note

Se usi Hive 3 e riscontri troppe connessioni al metastore di Hive, configura il parametro `datanucleus.connectionPool.maxPoolSize` per avere un valore inferiore o aumentare il numero di connessioni che il server di database è in grado di gestire. L'aumento del numero di connessioni è dovuto al modo in cui Hive calcola il numero massimo di connessioni JDBC. Per calcolare il valore ottimale per le prestazioni, consulta [Proprietà di configurazione di Hive](#).

Utilizzo di AWS Glue Data Catalog come metastore Hive

Con Amazon EMR rilascio 5.8.0 o successivi, è possibile configurare Hive per utilizzare AWS Glue Data Catalog come metastore. Si consiglia questa configurazione quando si richiede un metastore persistente o un metastore condiviso da diversi cluster, servizi e applicazioni o account AWS.

AWS Glue è un servizio di estrazione, trasformazione e caricamento (ETL) completamente gestito che consente di catalogare, filtrare e migliorare la qualità dei dati in modo semplice e conveniente, oltre a trasferirli in modo sicuro tra datastore. AWS Glue Data Catalog offre un repository di metadati unificato per varie origini dati e formati di dati, integrandosi con Amazon EMR così come con Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena e qualsiasi applicazione compatibile con il

metastore Apache Hive. AWS I crawler di Glue possono dedurre automaticamente lo schema dai dati di origine in Amazon S3 e archiviare i metadati associati nel Data Catalog. Per ulteriori informazioni sul Data Catalog, consulta la sezione relativa al [Popolamento di AWS Glue Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

AWS Glue è soggetto a costi separati. Viene applicata una tariffa mensile per l'archiviazione e l'accesso ai metadati nel Data Catalog, una tariffa oraria fatturata al minuto per i processi ETL di AWS Glue e per il runtime del crawler e, infine, una tariffa oraria fatturata al minuto per ciascun endpoint di sviluppo di cui è stato effettuato il provisioning. Il Data Catalog consente di archiviare fino a un milione di oggetti senza costi aggiuntivi. Se intendi archiviare più di un milione di oggetti, ti verrà addebitato 1 USD per ogni 100.000 oggetti oltre il milione. Un oggetto nel Data Catalog può essere una tabella, una partizione o un database. Per ulteriori informazioni, consulta [Prezzi di Glue](#).

Important

Se hai creato tabelle utilizzando Amazon Athena o Amazon Redshift Spectrum prima del 14 agosto 2017, i database e le tabelle sono archiviati in un catalogo gestito da Athena, che è separato da AWS Glue Data Catalog. Per integrare Amazon EMR con queste tabelle, dovrai eseguire l'aggiornamento a AWS Glue Data Catalog. Per ulteriori informazioni, consulta [Aggiornamento a AWS Glue Data Catalog](#) nella Guida per l'utente di Amazon Athena.

Specifiche di AWS Glue Data Catalog come metastore

È possibile specificare AWS Glue Data Catalog come metastore utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Quando si utilizza la CLI o l'API, è possibile utilizzare la classificazione di configurazione per Hive per specificare un Data Catalog. Inoltre, con Amazon EMR 5.16.0 e versioni successive, è possibile utilizzare la classificazione di configurazione per specificare un Catalogo di dati in un Account AWS diverso. Quando utilizzi la console, puoi specificare il Data Catalog utilizzando Advanced Options (Opzioni avanzate) o Quick Options (Opzioni rapide).

New console

Specifiche di AWS Glue Data Catalog come metastore Hive con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.

2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Application bundle, (Bundle dell'applicazione), scegli Core Hadoop, HBase o Custom (Personalizzato). Se personalizzi il cluster, assicurati di selezionare Hive o HCatalog come una delle tue applicazioni.
4. In Impostazioni Catalogo dati AWS Glue, seleziona la casella di controllo Utilizza per i metadati delle tabelle Hive.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di AWS Glue Data Catalog come metastore Hive con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Per Release (Versione), scegliere emr-5.8.0 o versioni successive.
4. In Release (Versione), selezionare Hive o HCatalog.
5. In AWS Glue Data Catalog settings (Impostazioni Glue Data Catalog), seleziona Use for Hive table metadata (Utilizza per metadati tabella Hive).
6. Scegliere altre opzioni per il cluster come appropriato, scegliere Next (Avanti), quindi configurare altre opzioni di cluster come appropriato per l'applicazione.

CLI

Specifica di AWS Glue Data Catalog come metastore Hive con la AWS CLI

Per ulteriori informazioni su come specificare una classificazione di configurazione utilizzando l'AWS CLI e l'API EMR, consulta [Configurazione delle applicazioni](#).

- Specificare il valore per `hive.metastore.client.factory.class` utilizzando la classificazione di configurazione `hive-site` come mostrato nell'esempio seguente.

```
[
```

```
{
  "Classification": "hive-site",
  "Properties": {
    "hive.metastore.client.factory.class":
    "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
  }
}
```

Nel rilascio EMR versioni 5.28.0, 5.28.1, 5.29.0 o 6.x, se stai creando un cluster utilizzando AWS Glue Data Catalog come metastore, imposta `hive.metastore.schema.validation` su `false`. Ciò impedisce a Hive e HCatalog di convalidare lo schema del metastore rispetto a MySQL. Senza questa configurazione, il gruppo di istanze primarie verrà sospeso dopo la riconfigurazione su Hive o HCatalog.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false"
    }
  }
]
```

Se disponi già di un cluster nelle versioni di rilascio EMR 5.28.0, 5.28.1 o 5.29.0, puoi impostare il gruppo di istanze primarie `hive.metastore.schema.validation` su `false` con le seguenti informazioni:

```
Classification = hive-site
Property       = hive.metastore.schema.validation
Value          = false
```

Per specificare un Data Catalog in un altro account AWS, aggiungi la proprietà `hive.metastore.glue.catalogid` come illustrato nel seguente esempio. Sostituisci *acct-id* con l'account AWS di 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"
    }
  }
]
```

Autorizzazioni IAM

Il profilo dell'istanza EC2 per un cluster deve disporre di autorizzazioni IAM per le operazioni AWS Glue. Inoltre, se abiliti la crittografia per oggetti AWS Glue Data Catalog, al ruolo deve essere consentito anche crittografare, decrittare e generare la AWS KMS key utilizzata per la crittografia dei dati.

Autorizzazioni per operazioni AWS Glue

Se utilizzi il profilo dell'istanza EC2 di default per Amazon EMR, non è richiesta alcuna operazione. La policy gestita `AmazonElasticMapReduceforEC2Role` collegata al `EMR_EC2_DefaultRole` consente tutte le operazioni AWS Glue necessarie. Tuttavia, se specifichi autorizzazioni e un profilo dell'istanza EC2 personalizzati, devi configurare le operazioni AWS Glue appropriate. Utilizza la policy `AmazonElasticMapReduceforEC2Role` gestita come punto di partenza. Per ulteriori informazioni, consulta [Ruolo di servizio per le istanze EC2 del cluster \(profilo dell'istanza EC2\)](#) nella Guida alla gestione di Amazon EMR.

Autorizzazioni per crittografare e decrittografare AWS Glue Data Catalog

Il tuo profilo dell'istanza ha bisogno dell'autorizzazione per crittografare e decrittare i dati utilizzando la tua chiave. Non è necessario configurare queste autorizzazioni se si applicano entrambe le seguenti istruzioni:

- Per abilitare la crittografia per oggetti AWS Glue Data Catalog utilizzare chiavi gestite per AWS Glue.
- Si utilizza un cluster che è nello stesso Account AWS come il AWS Glue Data Catalog.

In caso contrario, devi aggiungere la istruzione seguente alla policy delle autorizzazioni collegata al profilo dell'istanza EC2.

```
[
  {
    "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"
      }
    ]
  }
]
```

Per ulteriori informazioni sulla crittografia di AWS Glue Data Catalog, consulta la sezione [Crittografia del proprio Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

Autorizzazioni basate su risorse

Se utilizzi AWS Glue in combinazione con Hive, Spark o Presto in Amazon EMR, AWS Glue supporta policy basate su risorse per controllare l'accesso alle risorse del Data Catalog. Queste risorse includono tabelle, database, connessioni e funzioni definite dall'utente. Per ulteriori informazioni, consulta [Policy sulle risorse AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.

Quando utilizzi policy basate su risorse per limitare l'accesso ad AWS Glue da Amazon EMR, l'entità principale specificata nella policy delle autorizzazioni deve essere il ruolo ARN associato al profilo di istanza EC2 specificato al momento della creazione di un cluster. Ad esempio, per una policy basata sulle risorse collegata a un catalogo, è possibile specificare il ruolo ARN per il ruolo di servizio

predefinito per le istanze del cluster EC2, *EMR_EC2_DefaultRole*, come `Principal` utilizzando il formato mostrato nell'esempio seguente:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

L'*acct-id* può essere diverso dall'ID dell'account AWS Glue. Ciò consente l'accesso da cluster EMR in account diversi. Puoi specificare più entità principali, ognuna da un account diverso.

Considerazioni sull'utilizzo di AWS Glue Data Catalog

Considera i seguenti elementi durante l'utilizzo di AWS Glue Data Catalog come un metastore con Hive:

- L'aggiunta di JAR ausiliari tramite la shell Hive non è supportata. Per risolvere il problema, è possibile usare la `hive-site` classificazione di configurazione per impostare la proprietà `hive.aux.jars.path`, che aggiunge JAR ausiliarie percorso della classe Hive.
- Le [transazioni Hive](#) non sono supportate.
- La ridenominazione delle tabelle in AWS Glue non è supportata.
- Quando si crea una tabella Hive senza specificare una `LOCATION`, i dati nella tabella vengono archiviati nella posizione specificata dalla proprietà `hive.metastore.warehouse.dir`. Per impostazione predefinita, questa è una posizione in HDFS. Se un altro cluster necessita di accedere alla tabella, non ci riuscirà a meno che non disponga di autorizzazioni adeguate per il cluster che ha creato la tabella. Inoltre, poiché l'archiviazione HDFS è transitoria, se il cluster viene terminato, i dati della tabella vanno persi e la tabella deve essere ricreata. Ti consigliamo di specificare una `LOCATION` in Amazon S3 quando crei una tabella Hive utilizzando AWS Glue. In alternativa, puoi utilizzare la classificazione di configurazione `hive-site` per specificare una posizione in Amazon S3 per `hive.metastore.warehouse.dir`, che si applica a tutte le tabelle Hive. Se una tabella viene creata in una posizione HDFS e il cluster che l'ha creata è ancora in esecuzione, puoi aggiornare la posizione della tabella ad Amazon S3 in AWS Glue. Per ulteriori informazioni, consulta [Utilizzo di tabelle sulla console AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.
- I valori delle partizioni contenenti virgolette e apostrofi non sono supportati, ad esempio `PARTITION (owner="Doe 's")..`
- Le [statistiche di colonna](#) sono supportate per `emr-5.31.0` e versioni successive.
- L'utilizzo dell'[autorizzazione Hive](#) non è supportato. In alternativa, puoi prendere in considerazione l'utilizzo di [Policy basate sulle risorse AWS Glue](#). Per ulteriori informazioni, consulta [Utilizzo delle policy basate sulle risorse per l'accesso di Amazon EMR a AWS Glue Data Catalog](#).

- I [vincoli Hive](#) non sono supportati.
- L'[ottimizzazione basata sui costi in Hive](#) non è supportata.
- L'impostazione di `hive.metastore.partition.inherit.table.properties` non è supportata.
- L'utilizzo delle seguenti costanti metastore non è supportato: `BUCKET_COUNT`, `BUCKET_FIELD_NAME`, `DDL_TIME`, `FIELD_TO_DIMENSION`, `FILE_INPUT_FORMAT`, `FILE_OUTPUT_FORMAT`, `HIVE_FILTER_FIELD_LAST_ACCESS`, `HIVE_FILTER_FIELD_OWNER`, `HIVE_FILTER_FIELD_PARAMS`, `IS_ARCHIVED`, `META_TABLE_COLUMNS`, `META_TABLE_COLUMN_TYPES`, `META_TABLE_DB`, `META_TABLE_LOCATION`, `META_TABLE_NAME`, `META_TABLE_PARTITION_COLUMNS`, `META_TABLE_SERDE`, `META_TABLE_STORAGE`, `ORIGINAL_LOCATION`.
- Quando si utilizza un'espressione predicato, i valori espliciti devono trovarsi sul lato destro dell'operatore di confronto, altrimenti le query potrebbero avere esito negativo.
 - Corretto: `SELECT * FROM mytable WHERE time > 11`
 - Errato: `SELECT * FROM mytable WHERE 11 > time`
- Amazon EMR versioni 5.32.0 e 6.3.0 e successive supportano l'utilizzo di funzioni definite dall'utente (UDF) nelle espressioni predicate. Quando usi versioni precedenti, le tue query potrebbero fallire a causa del modo in cui Hive tenta di ottimizzare la relativa esecuzione.
- Le [tabelle temporali](#) non sono supportate.
- Ti consigliamo di creare tabelle utilizzando le applicazioni tramite Amazon EMR anziché crearle direttamente utilizzando AWS Glue. La creazione di una tabella tramite AWS Glue può causare la mancanza di campi obbligatori e originare eccezioni di query.
- In EMR 5.20.0 o versioni successive, l'eliminazione delle partizioni parallele viene abilitata automaticamente per Spark e Hive quando AWS Glue Data Catalog viene utilizzato come metastore. Questa modifica riduce significativamente i tempi di pianificazione delle query eseguendo più richieste in parallelo per recuperare le partizioni. Il numero totale di segmenti che è possibile eseguire contemporaneamente è compreso tra 1 e 10. Il valore di default è 5, che è l'impostazione consigliata. Puoi modificare questa impostazione specificando la proprietà `aws.glue.partition.num.segments` nella classificazione di configurazione `hive-site`. Se si verifica una limitazione, è possibile disattivare la funzione modificando il valore su 1. Per ulteriori informazioni, consulta [Struttura del segmento AWS Glue](#).

Utilizzo di un database MySQL o Amazon Aurora esterno

Per utilizzare un database MySQL esterno o Amazon Aurora come metastore Hive, si sovrascrivono i valori di configurazione predefiniti per il metastore in Hive per specificare la posizione del database esterno, su un'istanza Amazon RDS MySQL o su un'istanza Amazon Aurora PostgreSQL.

Note

Hive non supporta né impedisce l'accesso in scrittura simultaneo alle tabelle dei metastore. Se si condividono le informazioni del metastore tra due cluster, è necessario assicurarsi di non scrivere contemporaneamente sulla stessa tabella del metastore, a meno che non si scriva su partizioni diverse della stessa tabella del metastore.

La procedura seguente mostra come sovrascrivere i valori di configurazione predefiniti per la posizione del metastore Hive e avviare un cluster utilizzando la posizione del metastore riconfigurato.

Per creare un metastore situato al di fuori del cluster EMR

1. Creare un database MySQL o Aurora. Se si utilizza PostgreSQL, è necessario configurarlo dopo aver eseguito il provisioning del cluster. Solo MySQL è supportato durante la creazione del cluster. Per informazioni sulle differenze tra Aurora MySQL e Aurora PostgreSQL, consulta [Panoramica di Amazon Aurora MySQL](#) e [Utilizzo di Amazon Aurora PostgreSQL](#). Per informazioni generali su come creare un database Amazon RDS, consulta <https://aws.amazon.com/rds/>.
2. Modificare i gruppi di sicurezza per autorizzare le connessioni JDBC tra il database e il gruppo di sicurezza ElasticMapReduce-Master. Per ulteriori informazioni su come modificare i gruppi di sicurezza per l'accesso, consulta [Utilizzo dei gruppi di sicurezza gestiti da Amazon EMR](#).
3. Impostare i valori di configurazione JDBC in `hive-site.xml`:

Important

Se si forniscono informazioni sensibili, come password, all'API di configurazione di Amazon EMR, queste informazioni vengono visualizzate per gli account che dispongono di autorizzazioni sufficienti. Se si teme che queste informazioni possano essere visualizzate ad altri utenti, creare il cluster con un account amministrativo e limitare gli altri utenti (utenti IAM o con credenziali delegate) ad accedere ai servizi

del cluster creando un ruolo che nega esplicitamente le autorizzazioni alla chiave API `elasticmapreduce:DescribeCluster`.

- a. Creare un file di configurazione denominato `hiveConfiguration.json` contenente modifiche apportate a `hive-site.xml` come mostrato nel seguente esempio.

Sostituisci *hostname* (nome host) con l'indirizzo DNS dell'istanza Amazon RDS che esegue il database, e *username* (nome utente) e *password* con le credenziali per il database. Per ulteriori informazioni sulla connessione a istanze database MySQL e Aurora, consulta [Connessione a un'istanza database che esegue il motore del database di MySQL](#) e [Connessione a un cluster Athena DB](#) nella Guida per l'utente di Amazon RDS. `javax.jdo.option.ConnectionURL` è la stringa di connessione JDBC per un metastore JDBC. `javax.jdo.option.ConnectionDriverName` è il nome della classe del driver per un metastore JDBC.

I driver MySQL JDBC sono installati da Amazon EMR.

La proprietà `value` non può contenere spazi o ritorni a capo. Deve essere visualizzato tutto su una riga.

```
[
  {
    "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. Fare riferimento al file `hiveConfiguration.json` quando si crea il cluster come illustrato nel seguente comando AWS CLI. In questo comando, il file è memorizzato in locale ed è anche possibile caricare il file da Amazon S3 e farvi riferimento, per esempio `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

```
aws emr create-cluster --release-label emr-5.36.1 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=Hive --configurations file://hiveConfiguration.json --use-
default-roles
```

4. Effettua la connessione al nodo primario del cluster.

Per informazioni su come connettersi al nodo master, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida per la gestione di Amazon EMR.

5. Crea le tabelle Hive specificando il percorso su Amazon S3 immettendo un comando simile al seguente:

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
  key int,
  value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Aggiungere lo script Hive al cluster in esecuzione.

Il cluster Hive viene eseguito utilizzando il metastore situato in Amazon RDS. Avviare tutti i cluster Hive aggiuntivi che condividono questo metastore specificando la posizione del metastore.

Utilizzo del driver JDBC Hive

Puoi utilizzare gli strumenti di Business Intelligence più diffusi, quali Microsoft Excel, MicroStrategy, QlikView e Tableau con Amazon EMR per esplorare e visualizzare i dati. Molti di questi strumenti richiedono un driver Java Database Connectivity (JDBC) o un driver Open Database Connectivity (ODBC). L'algoritmo Amazon EMR supporta sia la connettività della JDBC che della ODBC.

L'esempio seguente mostra come utilizzare SQL Workbench/J come client SQL per connettersi a un cluster Hive in Amazon EMR. Per ulteriori driver, consulta [Utilizzo degli strumenti di Business Intelligence con Amazon EMR](#).

Prima di installare e lavorare con SQL Workbench/J, scaricare il pacchetto e installare il driver. I driver inclusi nel pacchetto supportano le versioni Hive disponibili in Amazon EMR versioni di rilascio 4.0 e successive. Per informazioni dettagliate sulle note di rilascio e sulla documentazione, consulta la documentazione PDF inclusa nel pacchetto.

- Download del pacchetto di driver JDBC Hive più recente

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Versioni precedenti del driver JDBC Hive

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

Per installare e configurare SQL Workbench

1. Scaricare il client SQL Workbench/J per il sistema operativo da <http://www.sql-workbench.net/downloads.html>.
2. Installare SQL Workbench/J. Per ulteriori informazioni, consulta la pagina relativa all'[installazione e all'avvio di SQL Workbench/J](#) nella guida per l'utente di SQL Workbench/J.
3. Utenti Linux, Unix, Mac OS X: in una sessione terminale, creare un tunnel SSH per il nodo master del cluster usando il seguente comando. Sostituisci *master-public-dns-name* con il nome DNS pubblico del nodo master e *path-to-key-file* con il percorso e il nome del file della chiave privata Amazon EC2 (.pem).

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Utenti di Windows: in una sessione PuTTY è necessario creare un tunnel SSH per il nodo master del cluster (tramite inoltro porta locale) con 10000 per Source port (Porta di origine) e *master-public-dns-name*:10000 per Destination (Destinazione). Sostituire *master-public-dns-name* con il nome DNS pubblico del nodo master.

4. Aggiungere il driver JDBC a SQL Workbench.
 - a. Nella finestra di dialogo Select Connection Profile (Seleziona profilo di connessione), fare clic su Manage Drivers (Gestisci driver).

- b. Fare clic sull'icona Create a new entry (Crea nuova voce) (pagina bianca).
- c. Nel campo Nome digitare **Hive JDBC**.
- d. Per Library (Libreria), fare clic sull'icona Select the JAR file(s) (Seleziona file JAR).
- e. Accedere alla posizione contenente i driver estratti. Selezionare i driver inclusi nella versione del pacchetto di driver JDBC scaricata e fare clic su Open (Apri).

Ad esempio, il pacchetto di driver JDBC potrebbe includere i seguenti JAR.

```
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. Nella finestra di dialogo Please select one driver (Selezionare un driver), selezionare `com.amazon.hive.jdbc41.HS2Driver`, OK.
5. Quando si torna alla finestra di dialogo Manage Drivers (Gestisci driver), verificare che il campo Classname sia popolato e selezionare OK.
6. Quando si ritorna alla finestra di dialogo Select Connection Profile (Seleziona profilo di connessione), verificare che il campo Driver sia impostato su Hive JDBC (JDBC Hive) e fornire la seguente stringa di connessione nel campo URL: `jdbc:hive2://localhost:10000/default`.
7. Selezionare OK per connettersi. Una volta stabilita la connessione, i dettagli di connessione appaiono nella parte superiore della finestra Workbench SQL/J.

Per ulteriori informazioni sull'utilizzo di Hive e l'interfaccia JDBC, consulta [HiveClient](#) e [HiveJDBCInterface](#) nella documentazione di Apache Hive.

Miglioramento delle prestazioni Hive

Amazon EMR offre funzioni che aiutano a ottimizzare le prestazioni quando si usa Hive per eseguire query e leggere e scrivere i dati salvati in Amazon S3.

S3 Select può migliorare le prestazioni delle query per i file CSV e JSON in alcune applicazioni mediante il pushdown dell'elaborazione ad Amazon S3.

Il committer ottimizzato EMRFS S3 è un'alternativa alla classe [OutputCommitter](#) che elimina le operazioni di elenco e rinominazione per migliorare le prestazioni durante la scrittura dei file in Amazon S3 utilizzando EMRFS.

Argomenti

- [Abilitazione del committer ottimizzato EMRFS S3 di Hive](#)
- [Utilizzo di S3 Select con Hive per migliorare le prestazioni](#)
- [Ottimizzazione MSCK](#)

Abilitazione del committer ottimizzato EMRFS S3 di Hive

Il committer ottimizzato EMRFS S3 di Hive è un modo alternativo con cui EMR Hive scrive i file per le query di inserimento quando si utilizza EMRFS. Il committer elimina le operazioni di elenco e rinominazione eseguite su Amazon S3 e migliora le prestazioni dell'applicazione. La funzione è disponibile a partire da EMR 5.34 e EMR 6.5.

Abilitazione del committer

Se desideri abilitare EMR Hive in modo da utilizzare `HiveEMRFSOptimizedCommitter` per eseguire il commit dei dati come impostazione predefinita per tutte le tabelle gestite da Hive e le tabelle esterne, utilizza la configurazione `hive-site` seguente nei cluster EMR 6.5.0 o EMR 5.34.0.

```
[
  {
    "classification": "hive-site",
    "properties": {
      "hive.blobstore.use.output-committer": "true"
    }
  }
]
```

Note

Non attivare questa funzione quando `hive.exec.parallel` è impostato su `true`.

Limitazioni

Ai tag si applicano le seguenti limitazioni di base:

- L'abilitazione di Hive per unire automaticamente file di piccole dimensioni non è supportata. La logica di commit Hive predefinita verrà utilizzata anche quando il committer ottimizzato è abilitato.
- Le tabelle Hive ACID non sono supportate. La logica di commit Hive predefinita verrà utilizzata anche quando il committer ottimizzato è abilitato.
- Il sistema di denominazione dei file per i file scritti viene modificata da quella di Hive `<task_id>_<attempt_id>_<copy_n>` a `<task_id>_<attempt_id>_<copy_n>_<query_id>`. Ad esempio, un file denominato `s3://warehouse/table/partition=1/000000_0` verrà modificato in `s3://warehouse/table/partition=1/000000_0-hadoop_20210714130459_ba7c23ec-5695-4947-9d98-8a40ef759222-1`. L'`query_id` di seguito è una combinazione di nome utente, marca temporale e UUID.
- Quando le partizioni personalizzate sono su file system diversi (HDFS, S3), questa funzione viene disabilitata automaticamente. La logica di commit Hive predefinita verrà utilizzata quando abilitata.

Utilizzo di S3 Select con Hive per migliorare le prestazioni

Con Amazon EMR 5.18.0 e versioni successive, puoi utilizzare [S3 Select](#) con Hive su Amazon EMR. S3 Select consente alle applicazioni il recupero di un solo sottoinsieme di dati da un oggetto. Per Amazon EMR, l'attività di calcolo del filtro di set di dati di grandi dimensioni per l'elaborazione viene "trasferita" dal cluster ad Amazon S3; in tal modo è possibile migliorare le prestazioni in alcune applicazioni e ridurre la quantità di dati trasferiti tra Amazon EMR e Amazon S3.

S3 Select è supportato con tabelle Hive basate su file CSV e JSON e impostando la variabile di configurazione `s3select.filter` su `true` durante la sessione Hive. Per maggiori informazioni ed esempi, consulta [Specifica di S3 Select nel codice](#).

S3 Select è adatto alla mia applicazione?

È consigliabile eseguire analisi comparative delle applicazioni con e senza S3 Select per vedere se l'utilizzo di tale caratteristica può risultare idoneo per l'applicazione.

Per determinare se l'applicazione è idonea per l'uso di S3 Select, utilizza le seguenti linee guida:

- La query filtra più di metà del set di dati originale.
- I predicati del filtro della query utilizzano colonne che dispongono di un tipo di dati supportato da Amazon S3 Select. Per ulteriori informazioni, consulta [Tipi di dati](#) nella Guida per l'utente di Amazon Simple Storage Service.
- La connessione di rete tra Amazon S3 e il cluster Amazon EMR ha una buona velocità di trasferimento e larghezza di banda disponibile. Amazon S3 non comprime le risposte HTTP, pertanto è probabile che le dimensioni della risposta aumentino per file di input compressi.

Considerazioni e limitazioni

- La crittografia lato server con chiavi di crittografia fornite dal cliente (SSE-C) e la crittografia lato client di Amazon S3 non sono supportate.
- La proprietà `AllowQuotedRecordDelimiters` non è supportata. Se questa proprietà è specificata, la query ha esito negativo.
- Sono supportati solo i file CSV e JSON in formato UTF-8. I file CSV e JSON a più righe non sono supportati.
- Sono supportati solo i file non compressi, gzip o bzip2.
- I caratteri di commento nell'ultima riga non sono supportati.
- Le righe vuote alla fine di un file non vengono elaborate.
- Hive su Amazon EMR supporta i tipi di dati primitivi supportati da S3 Select. Per ulteriori informazioni, consulta [Tipi di dati](#) nella Guida per l'utente di Amazon Simple Storage Service.

Specifiche di S3 Select nel codice

Per utilizzare S3 Select nella tabella Hive, crea la tabella specificando `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` come nome della classe `INPUTFORMAT` e specifica un valore per la proprietà `s3select.format` utilizzando la clausola `TBLPROPERTIES`.

Per impostazione predefinita, S3 Select è disabilitato quando si eseguono query. Abilita S3 Select impostando `s3select.filter` su `true` nella sessione Hive, come indicato di seguito. Gli esempi riportati di seguito mostrano come specificare S3 Select quando si crea una tabella da file CSV e JSON sottostanti e quindi si esegue la query della tabella utilizzando una semplice istruzione `select`.

Example Istruzione CREATE TABLE (CREA TABELLA) per una tabella basata su CSV

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,  
  col3 boolean  
)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
STORED AS  
INPUTFORMAT  
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'  
OUTPUTFORMAT  
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'  
LOCATION 's3://path/to/mycsvfile/'  
TBLPROPERTIES (  
  "s3select.format" = "csv",  
  "s3select.headerInfo" = "ignore"  
);
```

Example Istruzione CREATE TABLE (CREA TABELLA) per una tabella basata su JSON

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,  
  col3 boolean  
)  
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'  
STORED AS  
INPUTFORMAT  
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'  
OUTPUTFORMAT  
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'  
LOCATION 's3://path/to/json/'  
TBLPROPERTIES (  
  "s3select.format" = "json"  
);
```


Example Istruzione SELECT TABLE (SELEZIONA TABELLA)

```
SET s3select.filter=true;  
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

Ottimizzazione MSCK

Hive archivia un elenco di partizioni per ogni tabella nel suo metastore. Tuttavia, quando le partizioni vengono aggiunte o rimosse direttamente dal file system, il metastore Hive non riconosce queste modifiche. Il [comando MSCK](#) aggiorna i metadati delle partizioni nel metastore Hive per le partizioni che sono state aggiunte o rimosse direttamente dal file system. La sintassi per il comando è:

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

Hive implementa questo comando come segue:

1. Hive recupera tutte le partizioni per la tabella dal metastore. Dall'elenco dei percorsi delle partizioni che non esistono nel file system crea quindi un elenco di partizioni da eliminare dal metastore.
2. Hive raccoglie i percorsi delle partizioni presenti nel file system, li confronta con l'elenco delle partizioni del metastore e genera un elenco di partizioni che devono essere aggiunte al metastore.
3. Hive aggiorna il metastore utilizzando la modalità ADD, DROP o SYNC.

Note

Quando ci sono numerose partizioni nel metastore, l'esecuzione del passaggio per verificare se una partizione non esiste nel file system richiede molto tempo perché la chiamata API `exists` del file system deve essere effettuata per ogni partizione.

In Amazon EMR 6.5.0, Hive ha introdotto un flag chiamato `hive.emr.optimize.msck.fs.check`. Se abilitato, questo flag fa sì che Hive controlli la presenza di una partizione dall'elenco dei percorsi delle partizioni del file system generato nel passaggio 2 precedente invece di effettuare chiamate API del file system. In Amazon EMR 6.8.0, Hive ha abilitato questa ottimizzazione per impostazione predefinita, eliminando la necessità di impostare il flag `hive.emr.optimize.msck.fs.check`.

Utilizzo di Hive Live Long and Process (LLAP)

Amazon EMR 6.0.0 supporta la funzionalità Live Long and Process (LLAP) per Hive. LLAP utilizza daemon persistenti con memorizzazione nella cache intelligente in memoria per migliorare le prestazioni delle query rispetto alla precedente modalità di esecuzione predefinita del container Tez.

I demoni LLAP Hive sono gestiti ed eseguiti come servizio YARN. Poiché un servizio YARN può essere considerato un'applicazione YARN a esecuzione prolungata, alcune risorse del cluster sono dedicate a Hive LLAP e non possono essere utilizzate per altri carichi di lavoro. Per ulteriori informazioni, vedere [LLAP](#) e [YARN Service API](#).

Abilitazione di Hive LLAP su Amazon EMR

Per abilitare Hive LLAP su Amazon EMR, occorre fornire la seguente configurazione quando si avvia un cluster.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Per impostazione predefinita, Amazon EMR assegna circa il 60% delle risorse YARN cluster ai daemon LLAP Hive. È possibile configurare la percentuale di risorsa YARN cluster allocata a Hive LLAP e il numero di nodi di attività e core da considerare per l'allocazione LLAP Hive.

Ad esempio, la seguente configurazione avvia Hive LLAP con tre demoni su tre nodi task o core e assegna il 40% della risorsa YARN dei tre nodi core o task ai demoni LLAP Hive.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

```

    }
  }
]
```

È possibile utilizzare le seguenti configurazioni `hive-site` nell'API di classificazione per sovrascrivere le impostazioni di risorse LLAP predefinite.

Proprietà	Descrizione
<code>hive.llap.daemon.yarn.container.mb</code>	Dimensione totale del container del daemon LLAP (in MB)
<code>hive.llap.daemon.memory.per.instance.mb</code>	La memoria totale utilizzata dagli esecutori nel container daemon LLAP (in MB)
<code>hive.llap.io.memory.size</code>	Dimensione della cache per Input/Output LLAP
<code>hive.llap.daemon.num.executors</code>	Numero di esecutori per daemon LLAP

Avvio manuale di Hive LLAP nel cluster

Tutte le dipendenze e le configurazioni utilizzate da LLAP vengono inserite nell'archivio tar LLAP come parte dell'avvio del cluster. Se LLAP è abilitato utilizzando "`hive.llap.enabled`": "`true`", si consiglia di utilizzare la riconfigurazione Amazon EMR per apportare modifiche alla configurazione LLAP.

In caso contrario, per eventuali modifiche manuali a `hive-site.xml`, è necessario ricostruire l'archivio tar LLAP utilizzando il comando `hive --service llap`, come illustrato nell'esempio seguente.

```

# 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
```

Controllo dello stato di Hive LLAP

Utilizzare il comando seguente per verificare lo stato di Hive LLAP tramite Hive.

```
hive --service llapstatus
```

Utilizzare il seguente comando per verificare lo stato di Hive LLAP utilizzando YARN.

```
yarn app -status (name-of-llap-service)

# example:
yarn app -status llap0 | jq
```

Avvio o arresto di Hive LLAP

Poiché Hive LLAP viene eseguito come servizio YARN persistente, per arrestare o riavviare Hive LLAP occorre arrestare o riavviare il servizio YARN. I comandi seguenti illustrano come funziona.

```
yarn app -stop llap0
yarn app -start llap0
```

Ridimensionamento del numero di daemon di Hive LLAP

Utilizzare il comando seguente per ridurre il numero di istanze LLAP.

```
yarn app -flex llap0 -component llap -1
```

Per ulteriori informazioni, vedere [Flex di un componente di un servizio](#).

Crittografia in Hive

In questa sezione sono descritti i tipi di crittografia supportati da Amazon EMR.

Crittografia modulare Parquet in Hive

La crittografia modulare Parquet fornisce il controllo e la crittografia degli accessi a livello di colonna per migliorare la privacy e l'integrità dei dati archiviati nel formato di file Parquet. Questa funzionalità è disponibile in Amazon EMR Hive a partire dal rilascio 6.6.0.

Le soluzioni di sicurezza e integrità supportate in precedenza, che includono la crittografia dei file o la crittografia del livello di archiviazione, sono descritte in [Opzioni di crittografia](#) nella Guida alla gestione di Amazon EMR. Queste soluzioni possono essere utilizzate per i file Parquet, ma sfruttando le nuove funzionalità del meccanismo di crittografia integrato Parquet fornisce un accesso granulare a livello di colonna, oltre a miglioramenti in termini di prestazioni e sicurezza. Scopri di più su questa funzionalità nella pagina github di Apache [Crittografia modulare Parquet](#).

Gli utenti trasmettono le configurazioni a lettori e scrittori Parquet utilizzando configurazioni Hadoop. Le configurazioni dettagliate per consentire agli utenti di configurare lettori e scrittori per abilitare la crittografia e attivare le funzionalità avanzate sono documentate in [PARQUET-1854: Properties-driven Interface to Parquet Encryption Management](#)

Esempi di utilizzo

L'esempio seguente illustra la creazione e la scrittura su una tabella Hive che utilizza AWS KMS per la gestione delle chiavi di crittografia.

1. Implementa un KmsClient per il servizio AWS KMS come descritto nel documento [PARQUET-1373: Encryption Key Management Tools](#). L'esempio seguente mostra un frammento di implementazione.

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
```

```
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS_AWSKMS_CLIENT = AWSKMSClientBuilder
        .standard()
        .withRegion(Regions.US_WEST_2)
        .build();

    public static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
    private Configuration hadoopConfiguration;

    @Override
    public void initialize(Configuration configuration, String kmsInstanceID, String
kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
        hadoopConfiguration = configuration;
        kmsToken = accessToken;
    }

    @Override
    public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
KeyAccessDeniedException {
        String value = null;
        try {
            ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

            EncryptRequest req = new
EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
            ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

            byte[] base64EncodedValue = Base64.encode(ciphertext.array());
            value = new String(base64EncodedValue, Charset.forName("UTF-8"));
        } catch (AmazonClientException ae) {
```

```

        throw new KeyAccessDeniedException(ae.getMessage());
    }
    return value;
}

@Override
public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
KeyAccessDeniedException {
    byte[] arr = null;
    try {
        ByteBuffer ciphertext =
ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
        DecryptRequest request = new
DecryptRequest().withKeyId(masterKeyIdentifier).withCiphertextBlob(ciphertext);
        ByteBuffer decipheredtext =
AWSKMS_CLIENT.decrypt(request).getPlaintext();
        arr = new byte[decipheredtext.remaining()];
        decipheredtext.get(arr);
    } catch (AmazonClientException ae) {
        throw new KeyAccessDeniedException(ae.getMessage());
    }
    return arr;
}
}

```

2. Crea le tue chiavi di crittografia AWS KMS per il piè di pagina e le colonne con i tuoi ruoli IAM che hanno accesso, come descritto in [Creazione di chiavi](#) nella Guida per gli sviluppatori di AWS Key Management Service. Il ruolo IAM predefinito è EMR_ECS_default.
3. Nell'applicazione Hive su un cluster Amazon EMR, aggiungi il client precedente utilizzando l'istruzione ADD JAR, come descritto in [Risorse Apache Hive](#). Di seguito è mostrato un esempio di istruzione della policy:

```
ADD JAR 's3://location-to-custom-jar';
```

Un metodo alternativo consiste nell'aggiungere il JAR ad `auxlib` of Hive utilizzando un'azione bootstrap. Di seguito è riportato un esempio di riga da aggiungere all'azione bootstrap:

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

4. Imposta le seguenti configurazioni:

```
set
  parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
  parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;
```

5. Crea una tabella Hive in formato Parquet e specifica le chiavi AWS KMS in SERDEPROPERTIES e inserisci alcuni dati:

```
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. Verifica che quando crei una tabella esterna nella stessa posizione senza accesso alle chiavi AWS KMS (ad esempio, accesso al ruolo IAM negato), non sia possibile leggere i dati.

```
CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

SELECT * FROM ext_table;
```

7. L'ultima istruzione dovrebbe generare la seguente eccezione:

```
Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied
```


Crittografia in transito in HiveServer2

A partire da Amazon EMR versione 6.9.0, HiveServer2 (HS2) supporta TLS/SSL come parte della configurazione di sicurezza [Crittografia in transito in HiveServer2](#). Ciò influisce sul modo in cui ci si connette a HS2 in esecuzione su un cluster Amazon EMR con crittografia in transito abilitata. Per connettersi a HS2, è necessario modificare i valori dei parametri TRUSTSTORE_PATH e TRUSTSTORE_PASSWORD nell'URL di JDBC. L'URL seguente è un esempio di connessione JDBC per HS2 con i parametri richiesti:

```
jdbc:hive2://HOST_NAME:10000/  
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Utilizza le istruzioni appropriate per la crittografia HiveServer2 all'interno o all'esterno del cluster riportate di seguito.

On-cluster HS2 access

Se accedi a HiveServer2 utilizzando il client Beeline dopo avere effettuato l'accesso SSH al nodo primario, fai riferimento a `/etc/hadoop/conf/ssl-server.xml` per trovare i valori dei parametri TRUSTSTORE_PATH e TRUSTSTORE_PASSWORD utilizzando `ssl.server.truststore.location` e `ssl.server.truststore.password` della configurazione.

I seguenti comandi di esempio possono aiutarti a recuperare queste configurazioni:

```
TRUSTSTORE_PATH=$(sed -n '/ssl.server.truststore.location/,+2p' /etc/hadoop/conf/  
ssl-server.xml | awk -F "[><]" '/value/{print $3}')  
TRUSTSTORE_PASSWORD=$(sed -n '/ssl.server.truststore.password/,+2p' /etc/hadoop/  
conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
```

Off-cluster HS2 access

Se accedi a HiveServer2 da un client esterno al cluster Amazon EMR, puoi utilizzare uno dei seguenti approcci per recuperare i valori di TRUSTSTORE_PATH e TRUSTSTORE_PASSWORD:

- Converti il file PEM creato durante la [configurazione di sicurezza](#) in un file JKS e utilizza lo stesso nell'URL della connessione JDBC. Ad esempio, con openssl e keytool, utilizza i comandi seguenti:

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out
trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12
-destkeystore trustedCertificates.jks
```

- In alternativa, fai riferimento a `/etc/hadoop/conf/ssl-server.xml` per trovare i valori dei parametri `TRUSTSTORE_PATH` e `TRUSTSTORE_PASSWORD` utilizzando `ssl.server.truststore.location` e `ssl.server.truststore.password` della configurazione. Scarica il file `truststore` sul computer client e utilizza il percorso sul computer client come `TRUSTSTORE_PATH`.

Per ulteriori informazioni sull'accesso alle applicazioni da un client esterno al cluster Amazon EMR, consulta la sezione [Utilizzo del driver JDBC Hive](#).

Cronologia dei rilasci di Hive

La tabella seguente elenca la versione di Hive inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x o Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Hive

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server,

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
		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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.10.1	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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.7.0	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-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.1.0	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-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.14.1	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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.13.1	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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.2	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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.2	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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.4.1	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-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplib-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplib-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.3.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.2.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Etichetta di rilascio di Amazon EMR	Versione di Hive	Componenti installati con Hive
emr-4.1.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Note sul rilascio di Hive per versione

Argomenti

- [Amazon EMR 6.14.0: note di rilascio per Hive](#)
- [Amazon EMR 6.13.0: note di rilascio per Hive](#)
- [Amazon EMR 6.12.0: note di rilascio per Hive](#)
- [Amazon EMR 6.11.0: note di rilascio per Hive](#)
- [Amazon EMR 6.10.0: note di rilascio per Hive](#)
- [Amazon EMR 6.9.0 - Note sul rilascio per Hive](#)

- [Amazon EMR 6.8.0 - Note sul rilascio di Hive](#)
- [Amazon EMR 6.7.0 - Note sul rilascio di Hive](#)
- [Amazon EMR 6.6.0 - Note sul rilascio di Hive](#)

Amazon EMR 6.14.0: note di rilascio per Hive

Amazon EMR 6.14.0: modifiche di Hive

Tipo	Descrizione
Miglioramenti	HIVE-26762 : rimuovere la potatura degli operandi in <code>HiveFilterSetOpTransposeRule</code>
Correzioni di bug	HIVE-27582 : non memorizza nella cache il formato di input della tabella HBase in <code>FetchOperator</code>
Correzioni di bug	HIVE-26452 : NPE durante la conversione di JOIN in MAPJOIN e colonna JOIN a cui si fa riferimento più di una volta
Correzioni di bug	HIVE-26416 : <code>AcidUtils.isRawFormatFile()</code> lancia <code>InvalidProtocolBufferException</code> per il file non-ORC
Correzioni di bug	HIVE-26105 : mostra colonne mostra valori aggiuntivi se la colonna Commenti contiene caratteri cinesi specifici
Correzioni di bug	HIVE-25864 : l'ottimizzazione delle query di Hive crea un piano errato per il pushdown dei predicati con la funzione windowing
Correzioni di bug	HIVE-25224 : istruzioni multiple INSERT che coinvolgono tabelle con <code>bucketing_versions</code> diversi danno luogo a un errore

Tipo	Descrizione
Correzioni di bug	HIVE-24151 : MultiDelimiterDe sposta i dati se le stringhe contengono caratteri non-ASCII
Correzioni di bug	HIVE-23606 : (LLAP) Ritardo nella pulizia di DirectByteBuffer per EncodedReaderImpl
Correzioni di bug	HIVE-22165 : la sincronizzazione introdotta da HIVE-14296 su SessionManager.closeSession causa un'elevata latenza in un server Hive occupato
Correzioni di bug	HIVE-21304 : rendi più robusto l'utilizzo della versione bucketing

Amazon EMR 6.13.0: note di rilascio per Hive

Amazon EMR 6.13.0: modifiche di Hive

Tipo	Descrizione
Miglioramenti	Aggiornamento degli script Python per supportare Python3
Miglioramenti	HIVE-27097 : miglioramento della strategia di nuovi tentativi client e server MetaStore
Correzioni di bug	HIVE-21778 : CBO: "Struct is not null" viene valutato come annullabile, causando sempre la mancanza di filtro nella query
Correzioni di bug	HIVE-21009 : aggiunta della possibilità per l'utente di impostare l'utente vincolato

Tipo	Descrizione
Correzioni di bug	HIVE-22661 : la compattazione ha esito negativo su una tabella non suddivisa in bucket con dati caricati nel percorso
Correzioni di bug	HIVE-19718 : l'aggiunta di partizioni in blocco consente inoltre di recuperare la tabella per ogni partizione
Correzioni di bug	HIVE-22173 : la query con viste laterali multiple si blocca durante la compilazione
Correzioni di bug	HIVE-27088 : risultati errati quando inner e outer join con i filtri post join vengono combinati
Correzioni di bug	HIVE-21935 : vettorizzazione Hive: prestazioni ridotte con UDF vettorizzate
Correzioni di bug	HIVE-25299 : il trasferimento del timestamp su tipi di dati numerici non è corretto per i fusi orari non UTC
Correzioni di bug	HIVE-24626 : LLAP: i thread dei lettori potrebbero esaurirsi se tutti i thread di I/O elevator sono occupati per essere accodati a un altro lettore con coda completa
Correzioni di bug	HIVE-27029 : la query hive ha esito negativo con l'errore di chiusura del file system, rielaborazione completata per HIVE-26352
Correzioni di bug	HIVE-26352 : il controllo dell'accesso alla coda Tez ha esito negativo con l'eccezione sulla compattazione GSS
Correzioni di bug	HIVE-24590 : la registrazione delle operazioni continua a perdere gli appender log4j

Tipo	Descrizione
Correzioni di bug	HIVE-24552 : possibili perdite o accumuli di connessioni HMS nelle partizioni loadDynam icPartitions
Correzioni di bug	HIVE-27069 : risultati errati con map join del bucket
Correzioni di bug	HIVE-27344 : aggiunta di un controllo nullo in RecordReaderImpl#close
Correzioni di bug	HIVE-27439 : spazio di supporto in decimale
Correzioni di bug	HIVE-27267 : risultati errati quando si esegue il map join dei bucket su una colonna decimale con sottoquery
Correzioni di bug	HIVE-21986 : interfaccia utente Web HiveServe r: impostazione di Strict-Transport-Security nell'intestazione della risposta predefinita
Correzioni di bug	HIVE-22148 : i token di delega S3A non vengono aggiunti nella configurazione del processo di Compactor.
Correzioni di bug	HIVE-22622 : Hive consente di creare una struttura con nomi di attributi duplicati
Correzioni di bug	HIVE-22008 : l'operatore LIKE deve corrispon dere a un input multilinea
Correzioni di bug	HIVE-23144 : LLAP: possibilità di QueryTracker di eseguire la pulizia su ServiceStop
Correzioni di bug	HIVE-22391 : NPE durante il controllo della cache dei risultati delle query di Hive

Tipo	Descrizione
Correzioni di bug	HIVE-23305 : NullPointerException in LlapTaskSchedulerService addNode a causa della condizione di gara
Correzioni di bug	HIVE-22178 : Parquet FilterPredicate genera CastException dopo SchemaEvolution
Correzioni di bug	HIVE-21517 : correzione di AggregateStatsCache
Correzioni di bug	HIVE-21825 : miglioramento del messaggio di errore del client quando l'HA attivo/passivo è abilitato
Correzioni di bug	HIVE-23389 : FilterMergeRule può causare AssertionError
Correzioni di bug	HIVE-22767 : Beeline non analizza correttamente il punto e virgola nei commenti
Correzioni di bug	HIVE-22996 : l'analisi di BasicStats dovrebbe verificare in modo proattivo la presenza di stringhe nulle o vuote
Correzioni di bug	HIVE-22808 : HiveRelFieldTrimmer non gestisce HiveTableFunctionScan
Correzioni di bug	HIVE-22437 : i metadati LLAP memorizzano NPE sul blocco dei metadati.
Correzioni di bug	HIVE-22606 : AvroSerde registra avro.schema.literal nel livello INFO
Correzioni di bug	HIVE-22713 : la propagazione costante non dovrebbe essere eseguita per la struttura Join-Fil(*)-RS

Tipo	Descrizione
Correzioni di bug	HIVE-21624 : LLAP: i parametri della CPU a livello di thread non funzionano
Correzioni di bug	HIVE-22815 : riduzione della creazione di oggetti non necessari nel file system in MROutput
Correzioni di bug	HIVE-23060 : query con errore "Grouping sets expression is not in GROUP BY key. Error encountered near token"
Correzioni di bug	HIVE-22236 : impossibile creare la vista selezionando la vista contenente la sottoquery NOT IN
Correzioni di bug	HIVE-19886 : i log possono essere indirizzati a 2 file se viene utilizzato -hiveconf hive.log.file
Correzioni di bug	HIVE-20620 : collisioni del manifesto durante l'inserimento in tabelle MM ordinate in bucket con partizionamento dinamico
Correzioni di bug	HIVE-14557 : Nullpointer quando sia SkewJoin che Mapjoin sono abilitati
Correzioni di bug	HIVE-20471 : problemi con l'ottenimento del percorso predefinito del database
Correzioni di bug	HIVE-20598 : correzione degli errori di battitura nei calcoli HiveAlgorithmsUtil
Correzioni di bug	HIVE-14737 : problema di accesso a /logs in un'interfaccia utente Web Kerberized Hive Server 2
Correzioni di bug	HIVE-20733 : GenericUDFOPEqualNS non può utilizzare = nelle descrizioni dei piani

Tipo	Descrizione
Correzioni di bug	HIVE-20848 : dopo aver impostato UpdateInp utAccessTimeHook, la query ha esito negativo con Table Not Found.
Correzioni di bug	HIVE-18929 : il metodo humanReadableInt in HiveStringUtils.java ha una condizione di gara.
Correzioni di bug	HIVE-20841 : LLAP: possibilità di configurazione delle porte dinamiche
Correzioni di bug	HIVE-20930 : VectorCoalesce in modalità FILTER non ha effetto
Correzioni di bug	HIVE-21007 : Semi join + Union può causare piani errati
Correzioni di bug	HIVE-21074 : lo sfoltimento delle query delle tabelle in bucket di Hive non funziona per la condizione IS NOT NULL
Correzioni di bug	HIVE-21223 : CachedStore restituisce una partizione nulla quando la partizione non esiste
Correzioni di bug	HIVE-19625 : NPE potenziale ed eccezione effettiva nascosta in Hive#copyFiles
Correzioni di bug	HIVE-17020 : la deduplicazione RS aggressiva può rimuovere erroneamente il ramo dell'albero OP
Correzioni di bug	HIVE-20168 : accesso di ReduceSinkOperator nascosto
Correzioni di bug	HIVE-20879 : l'utilizzo di un valore nullo in un'espressione di proiezione causa CastException

Tipo	Descrizione
Correzioni di bug	HIVE-20888 : TxnHandler: sort() richiamato su elenchi immutabili
Correzioni di bug	HIVE-19948 : HiveCli non divide correttamente il comando con punto e virgola se le virgolette sono all'interno della stringa
Correzioni di bug	HIVE-20621 : GetOperationStatus viene chiamato in resultset.next causando un rallentamento incrementale
Correzioni di bug	HIVE-20854 : impostazioni predefinite sensibili: l'intervallo heartbeat di Hive Zookeeper è di 20 minuti, modifica in 2
Correzioni di bug	HIVE-20330 : HCatLoader non è in grado di gestire più oggetti InputJobInfo per un processo con più input
Correzioni di bug	HIVE-20787 : il caso MapJoinBytesTableContainer dummyRow non gestisce il riutilizzo
Correzioni di bug	HIVE-20331 : la query con union all, lateral view e Join restituisce un errore "cannot find parent in the child operator"
Correzioni di bug	HIVE-19968 : l'eccezione UDF non viene eliminata
Correzioni di bug	HIVE-20410 : l'interruzione dell'operazione Insert Overwrite sulla tabella transazionale causa l'errore "Not enough history available for..."
Correzioni di bug	HIVE-20059 : lo streaming di Hive dovrebbe provare il prefisso shade incondizionatamente in caso di eccezione

Tipo	Descrizione
Correzioni di bug	HIVE-19424 : NPE in MetaDataFormatters
Correzioni di bug	HIVE-20355 : pulizia del parametro di HiveConnection.setSchema
Correzioni di bug	HIVE-20858 : Serializer non è inizializzato correttamente con la configurazione in Utilities.createEmptyBuckets
Correzioni di bug	HIVE-20424 : schematool non deve inquinare la cronologia di beeline
Correzioni di bug	HIVE-20338 : LLAP: obbligo dell'ID sintetico dei file per i file system che hanno protocolli HDFS con semantica di mutazione POSIX
Correzioni di bug	HIVE-11708 : gli operatori logici generano ClassCastExceptions con NULL
Correzioni di bug	HIVE-21082 : in HPL/SQL, l'istruzione declare non supporta variabili di tipo carattere
Correzioni di bug	HIVE-16690 : configurazione dell'edge del prodotto cartesiano Tez in base alla dimensione del cluster LLAP
Correzioni di bug	HIVE-21296 : l'eliminazione della partizione varchar genera un'eccezione
Correzioni di bug	HIVE-14516 : OrclInputFormat.SplitGenerator.callInternal
Correzioni di bug	HIVE-20981 : streaming/AbstractRecordWriter perde HeapMemoryMonitor

Tipo	Descrizione
Correzioni di bug	HIVE-20043 : HiveServer2: SessionState ha un blocco di sincronizzazione statico attorno a un AtomicBoolean
Correzioni di bug	HIVE-20191 : l'applicazione patch PreCommit viene eseguita correttamente se la patch è vuota
Correzioni di bug	HIVE-20400 : la creazione della tabella dovrebbe sempre utilizzare un percorso completo per evitare potenziali ambiguità FS
Correzioni di bug	Aggiunta del controllo nullo per skewedInfo prima di accedere alle colonne asimmetriche

Amazon EMR 6.12.0: note di rilascio per Hive

Amazon EMR 6.12.0: modifiche di Hive

Tipo	Descrizione
Miglioramenti	Aggiunto il supporto per JDK 11 e JDK 17 Runtime
Miglioramenti	Aggiunta del supporto per eseguire query sui nomi delle colonne di parole chiave riservate e con distinzione tra maiuscole e minuscole quando utilizzi S3 Select. Per utilizzarlo, definisci la proprietà della tabella nel formato "s3select.column.mapping" = " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ..."
Miglioramenti	HIVE-23133 : le operazioni numeriche possono avere risultati diversi tra gli archi hardware

Tipo	Descrizione
Miglioramenti	HIVE-27145 : utilizzo di StrictMath per le restanti funzioni matematiche come seguito di HIVE-23133
Correzioni di bug	Correzione dell'incompatibilità con i jolly nelle API HMS <code>get_partitions_by_filter</code> e <code>get_num_partitions_by_filter</code> causata dal porting di HIVE-22900 in EMR Hive 6.4.0
Correzioni di bug	HIVE-26736 : errore di autorizzazione per le viste annidate con clausola WITH
Correzioni di bug	HIVE-22416 : log delle operazioni relativi a MR mancanti quando è abilitata l'esecuzione parallela
Correzioni di bug	HIVE-19653 : pushdown errato dei predicati per groupby con set di raggruppamento
Correzioni di bug	HIVE-22094 : query con esito negativo con <code>ClassCastException: hive.ql.exec.vector.DecimalColumnVector non può essere lanciata su hive.ql.exec.vector.decimal64ColumnVector</code>
Correzioni di bug	HIVE-26340 : l'operatore PTF vettorizzato ha esito negativo se la query ha la funzione di finestra in maiuscolo
Correzioni di bug	HIVE-26184 : <code>COLLECT_SET</code> con <code>GROUP BY</code> è lento quando alcune chiavi sono molto asimmetriche
Correzioni di bug	HIVE-26373 : <code>ClassCastException</code> durante la lettura dei timestamp dalla tabella HBase con dati Avro

Tipo	Descrizione
Correzioni di bug	HIVE-26388 : ClassCastException quando non è presente una colonna di tipo stringa nella tabella di origine della query CTAS Aggiornamento di HIVE-26172 : Hive - Aggiornamento di Ant a 1.10.11 grazie a CVE-2021-36373 e CVE-2021-36374
Correzioni di bug	HIVE-26114 : la correzione della connessione jdbc hiveserver2 utilizzando il comando dfs con lo spazio del prefisso causa un'eccezione
Correzioni di bug	HIVE-26396 : la funzione trunc ha un problema con l'intercettazione di precisione e il risultato contiene 0
Correzioni di bug	HIVE-26446 : HiveProtoLoggingHook non è in grado di compilare il campo TablesWritten per le tabelle partizionate.
Correzioni di bug	HIVE-26639 : ConstantVectorExpression e ExplainTask non dovrebbero basarsi sul set di caratteri predefinito
Correzioni di bug	HIVE-22670 : ArrayIndexOutOfBoundsException quando si utilizza un lettore vettoriale per leggere un file parquet
Correzioni di bug	HIVE-23607 : problema di autorizzazione: la creazione di una vista su di un'altra viene eseguita correttamente ma la modifica della visualizzazione ha esito negativo
Correzioni di bug	HIVE-25498 : la query con più di 31 funzioni distinte di conteggio restituisce un risultato errato

Tipo	Descrizione
Correzioni di bug	HIVE-25780 : DistinctExpansion crea più di 64 set di raggruppamento II
Correzioni di bug	HIVE-23868 : specifiche della funzione di windowing: supporta 0 precedenti/seguenti
Correzioni di bug	HIVE-24539 : la generazione dello schema OrclInputFormat deve rispettare il delimitatore di colonna
Correzioni di bug	HIVE-23476 : LLAP: preallocazione delle arene anche per il caso mmap
Correzioni di bug	HIVE-25806 : possibile perdita di dati in LlapCacheAwareFs - Parquet, LLAP I/O
Correzioni di bug	HIVE-23498 : disabilitazione del metodo HTTP Trace su ThriftHttpCliService
Correzioni di bug	HIVE-25729 : ThriftUnionObjectInspector deve ricevere una notifica quando è completamente avviato
Correzioni di bug	HIVE-23846 : possibilità di evitare la serializzazione e la deserializzazione non necessarie dei bitvector
Correzioni di bug	HIVE-24233 : la sottoquery except genera nullpointer con cbo disabilitato
Correzioni di bug	HIVE-24276 : vulnerabilità scripting cross-site (XSS) di HiveServer2 loggerconf.jsp
Correzioni di bug	HIVE-25721 : il risultato di outer join è errato
Correzioni di bug	HIVE-25223 : selezione con limite non restituisce alcuna riga su una tabella non nativa

Tipo	Descrizione
Correzioni di bug	HIVE-25794 : CombineHiveRecordReader: le istruzioni di log in un ciclo causano una pressione della memoria
Correzioni di bug	HIVE-23602 : utilizzo di Java Concurrent Package per Operation Handle Set
Correzioni di bug	HIVE-24045 : nessun log relativo alla creazione del database predefinito
Correzioni di bug	HIVE-24305 : lo schema decimale avro non inserisce correttamente il dimensionamento o la precisione se il valore è racchiuso tra virgolette
Correzioni di bug	HIVE-25844 : errori di deserializzazione eccezionali possono causare la terminazione immediata di beeline
Correzioni di bug	HIVE-25040 : Drop database cascade non può rimuovere le funzioni persistenti
Correzioni di bug	HIVE-23501 : AOOB in VectorDeserializeRow quando i tipi complessi vengono convertiti in tipi primitivi
Correzioni di bug	HIVE-23704 : il server HTTP Thrift non gestisce correttamente l'handle di autenticazione
Correzioni di bug	HIVE-23529 : CTAS è interrotto per uniontype quando row_deserialize
Correzioni di bug	HIVE-24144 : getIdentifierQuoteString in HiveDatabaseMetadata restituisce un valore errato

Tipo	Descrizione
Correzioni di bug	HIVE-23850 : consente PPD quando il soggetto non è una colonna con set di raggruppamento presenti
Correzioni di bug	HIVE-24036 : Kryo Exception durante la serializzazione del piano per la chiamata UDF getSplits
Correzioni di bug	HIVE-25919 : ClassCastException quando si invia il predicato di colonna booleano in HBaseStorageHandler
Correzioni di bug	HIVE-25261 : RetryingHMSHandler dovrebbe racchiudere MetaException con una breve descrizione del target
Correzioni di bug	HIVE-24792 : potenziale perdita di thread durante il funzionamento
Correzioni di bug	HIVE-23409 : Se la riapertura dell'applicazione TezSession ha esito negativo a causa dell'interruzione del servizio Timeline, la TEZSession predefinita di SessionPool viene chiusa dopo un nuovo tentativo
Correzioni di bug	HIVE-23615 : nessuna deferenza a puntatori nulli nella classe dei comandi di Beeline
Correzioni di bug	HIVE-24849 : creazione di un timeout esterno per il socket della tabella quando la posizione ha un numero elevato di file (riguarda la versione 3.1.2)
Correzioni di bug	HIVE-24193 : la query di selezione sulla tabella hive acid rinominata non produce alcun output

Tipo	Descrizione
Correzioni di bug	HIVE-25209 : query SELECT con funzione SUM che produce risultati imprevisti
Correzioni di bug	HIVE-23666 : checkHashModeEfficiency viene saltato quando un operatore groupby non dispone di un set di raggruppamento
Correzioni di bug	HIVE-23873 : la query della tabella JDBCStorageHandler di Hive ha esito negativo con NPE quando CBO è disattivato
Correzioni di bug	HIVE-24149 : HiveStreamingConnection non chiude la connessione HMS
Correzioni di bug	HIVE-25561 : l'attività interrotta non deve salvare il file. (Riguarda le versioni 2.x e 3.x)
Correzioni di bug	HIVE-25683 : chiusura del lettore in AcidUtils.isRawFormatFile
Correzioni di bug	HIVE-24294 : le sessioni TEZSessionPool possono generare AssertionError
Correzioni di bug	HIVE-24182 : problema di autorizzazione Ranger con UDF permanenti
Correzioni di bug	HIVE-22805 : la vettorizzazione con array o mappa condizionale non è implementata e genera un errore
Correzioni di bug	HIVE-22828 : Decimal64: le istruzioni NVL e CASE convertono implicitamente decimal64 in 128
Correzioni di bug	HIVE-21398 : le colonne che contengono o statistiche stimate non devono essere considerate chiavi univoche

Tipo	Descrizione
Correzioni di bug	HIVE-22490 : l'aggiunta di jar con caratteri speciali nel loro percorso genera un errore
Correzioni di bug	HIVE-22700 : le compattazioni possono perdere memoria se non autorizzate
Correzioni di bug	HIVE-22053 : il nome della funzione non è normalizzato durante la creazione della funzione
Correzioni di bug	HIVE-22595 : gli inserimenti dinamici delle partizioni hanno esito negativo sulla tabella Avro con schema esterno
Correzioni di bug	HIVE-21795 : la riga di riepilogo del rollup potrebbe mancare quando si verifica un mapjoin su una tabella partizionata
Correzioni di bug	HIVE-22987 : ClassCastException in VectorCoalesce quando DataTypePhysicalVariation è nullo
Correzioni di bug	HIVE-22814 : ArrayIndexOutOfBoundsException nella vettorizzazione getDataTypePhysicalVariation
Correzioni di bug	HIVE-22523 : il gestore degli errori in LlapRecordReader potrebbe bloccarsi se la coda è piena
Correzioni di bug	HIVE-21796 : ArrayWritableObjectInspector.equals può richiedere del tempo $O(2^{\text{nesting_depth}})$
Correzioni di bug	HIVE-22929 : prestazioni: l'analisi degli identificatori tra virgolette utilizza Regex usa e getta tramite String.replaceAll()

Tipo	Descrizione
Correzioni di bug	HIVE-21641 : il client esterno Llap restituisce colonne decimali con precisione/dimensionamento diversi rispetto a beeline
Correzioni di bug	HIVE-22207 : Tez: SplitGenerator genera NumberFormatException quando "dfs.block.size" sul cluster è "128m"
Correzioni di bug	HIVE-22114 : la query di inserimento per la sola tabella partizionata ha esito negativo quando tutti i bucket sono vuoti
Correzioni di bug	HIVE-22165 : la sincronizzazione introdotta da HIVE-14296 su SessionManager.closeSession causa un'elevata latenza in un server hive occupato
Correzioni di bug	HIVE-22744 : TezTask per il vertice con più di un outedge dovrebbe avere una memoria di ordinamento proporzionale
Correzioni di bug	HIVE-22072 : l'alterazione della tabella per apportare una modifica alla colonna non aggiorna i riferimenti ai vincoli
Correzioni di bug	HIVE-22075 : correzione della regressione max-reducers=1 da HIVE-14200
Correzioni di bug	HIVE-22527 : Hive su Tez: il processo di unione di file di piccole dimensioni viene inviato in un'altra coda (coda predefinita)
Correzioni di bug	HIVE-22816 : QueryCache: le query che utilizzano le viste possono averle memorizzate nella cache dopo l'espansione CTE

Tipo	Descrizione
Correzioni di bug	HIVE-22733 : dopo aver disabilitato la proprietà del log delle operazioni in hive, HS2 sta ancora salvando il log delle operazioni
Correzioni di bug	HIVE-22699 : le UDF della maschera devono mascherare il valore numerico 0
Correzioni di bug	HIVE-23356 : l'aggregazione di hash è sempre disabilitata durante l'elaborazione di query con espressioni di set di raggruppamento.
Correzioni di bug	HIVE-21568 : HiveRelOptUtil.isRowFilteringPlan dovrebbe saltare il progetto
Correzioni di bug	HIVE-21760 : l'ottimizzazione del lavoro condiviso deve essere ignorata per i join SMB
Correzioni di bug	HIVE-22712 : il driver ReExec esegue l'invio della query nella coda predefinita indipendentemente dalla coda definita dall'utente
Correzioni di bug	HIVE-21397 : la tabella BloomFilter per hive Managed [ACID] non funziona come previsto
Correzioni di bug	HIVE-23011 : lo strumento di ottimizzazione del lavoro condiviso dovrebbe controllare i predicati residui quando si confrontano i join
Correzioni di bug	HIVE-21412 : PostExecOrcFileDump non funziona con le tabelle ACID
Correzioni di bug	HIVE-22201 : ConvertJoinMapJoin#checkShuffleSizeForLargeTable genera ArrayIndexOutOfBoundsException se non è selezionata alcuna tabella grande

Tipo	Descrizione
Correzioni di bug	HIVE-21971 : HS2 perde il classloader a causa di "ReflectionUtils::CONSTRUCTOR_CACHE" con funzioni temporanee + GenericUDF
Correzioni di bug	HIVE-21368 : vettorizzazione: conversione Decimal64 -> HiveDecimal non necessaria
Correzioni di bug	HIVE-25416 : perdita di memoria del metastore di Hive a causa del bug datanucleus-api-jdo
Correzioni di bug	HIVE-22219 : la disattivazione di un gestore di nodi blocca il riavvio del servizio LLAP
Correzioni di bug	HIVE-21793 : CBO recupera le statistiche delle colonne anche se hive.stats.fetch.column.stats è impostato su falso
Correzioni di bug	HIVE-22163 : CBO: l'abilitazione di CBO attiva la stima delle statistiche, anche quando la stima è disabilitata
Correzioni di bug	HIVE-18735 : creazione di una tabella come se si verificasse la perdita dell'attributo transazionale
Correzioni di bug	HIVE-22433 : Hive JDBC Storage Handler: risultati errati recuperati da BOOLEAN e TIMESTAMP DataType dall'origine dati JDBC
Correzioni di bug	HIVE-19430 : ObjectStore.cleanNotificati onEvents OutOfMemory su un gran numero di eventi in sospeso
Correzioni di bug	HIVE-20785 : nome chiave errato nel metodo JDBC DatabaseMetadata.getPrimaryKeys

Tipo	Descrizione
Correzioni di bug	HIVE-16116 : Beeline genera NPE quando <code>beeline.hiveconfvariables={}</code> in <code>beeline.properties</code>
Correzioni di bug	HIVE-20066 : <code>hive.load.data.owner</code> viene confrontato con il principale completo
Correzioni di bug	HIVE-20489 : spiegazione del piano dei blocchi delle query
Correzioni di bug	HIVE-21033 : dimenticare di chiudere l'operazione interrompe l'ulteriore output di HiveServer2
Correzioni di bug	HIVE-19888 : avviso fuorviante "METASTORE_FILTER_HOOK will be ignored" da SessionState
Correzioni di bug	HIVE-20303 : INSERT OVERWRITE TABLE db.table PARTITION (...) IF NOT EXISTS genera InvalidTableException
Correzioni di bug	HIVE-16144 : CompactionInfo non ha equals/hashCode ma viene utilizzato in Set
Correzioni di bug	HIVE-20818 : le viste create con una sottoquery e WHERE considerano le viste a cui si fa riferimento nella sottoquery come input diretto
Correzioni di bug	HIVE-21005 : LLAP: la lettura di più strisce per split perde ZlibCodecs
Correzioni di bug	HIVE-20771 : LazyBinarySerde ha esito negativo su strutture vuote.
Correzioni di bug	HIVE-18852 : messaggio di errore fuorviante nella convalida della tabella alter

Tipo	Descrizione
Correzioni di bug	HIVE-21124 : HPL/SQL non supporta l'istruzione CREATE TABLE LIKE
Correzioni di bug	HIVE-20935 : il caricamento del file tarball del pacchetto llap ha esito negativo in EC2, causando un errore di avvio del servizio LLAP
Correzioni di bug	HIVE-20409 : Hive ACID: Aggiorna/elimina/unionisci non pulisce la directory di staging hdfs
Correzioni di bug	HIVE-20570 : union ALL con hive.optimize.union.remove=true contiene un piano errato
Correzioni di bug	HIVE-20421 : entità di caratteri non valida "\b" in hive-default.xml.template
Correzioni di bug	HIVE-19133 : i parametri delle prestazioni fase per fase di HS2 WebUI non vengono visualizzati correttamente
Correzioni di bug	HIVE-18977 : l'elenco delle partizioni restituisce risultati diversi con JDO e Direct SQL
Correzioni di bug	HIVE-20034 : ripristino delle modifiche alla gestione delle eccezioni di MetaStore per la compatibilità con le versioni precedenti
Correzioni di bug	HIVE-20672 : il thread di registrazione in LlapTaskSchedulerService dovrebbe riportare ogni intervallo fisso
Correzioni di bug	HIVE-12812 : abilitazione di mapred.input.dir.recursive per impostazione predefinita per supportare l'unione con la funzione aggregata
Correzioni di bug	HIVE-20147 : l'importazione dello streaming di Hive si basa sulla registrazione sincronizzata

Tipo	Descrizione
Correzioni di bug	HIVE-19203 : problema di sicurezza dei thread in HiveMetastore
Correzioni di bug	HIVE-20091 : Tez: aggiunta di credenziali di sicurezza per l'output di FileSinkOperator
Correzioni di bug	HIVE-16906 : Hive ATSHook dovrebbe verificar e la presenza di yarn.timeline-service.enabled prima di connettersi ad ATS
Correzioni di bug	HIVE-20714 : SHOW tblproperties per una proprietà singola restituisce il valore nella colonna del nome
Correzioni di bug	HIVE-24730 : le classi Shims sovrascrivono silenziosamente i valori di hive-site.xml e tez-site.xml
Correzioni di bug	HIVE-22055 : il conteggio delle selezioni fornisce un risultato errato dopo il caricamento dei dati dal file di testo

Amazon EMR 6.11.0: note di rilascio per Hive

Amazon EMR 6.11.0: modifiche di Hive

Tipo	Descrizione
Miglioramenti	Aggiunto il supporto per l'eliminazione multithread delle partizioni per migliorare le prestazioni di eliminazione di queste ultime
Miglioramenti	Supporto della lettura di file di query Hive codificati

Tipo	Descrizione
Miglioramenti	Abilitazione di Tez Shuffle Handler per impostazione predefinita per i processi Hive su Tez
Bug	Aggiunta un'opzione per abilitare la distribuzione deterministica delle chiavi sui riduttori per correggere il risultato errato quando <code>hive.groupby.skewindata</code> è abilitato (segnalato in HIVE-20220)
Bug	Correzione dell'errore di calcolo delle statistiche quando è configurato il nome della partizione e predefinita
Bug	Rispetto di tutti i parametri di classificazione SSL personalizzati passati quando SSL viene configurato immediatamente per HiveServer2 in un cluster con crittografia in transito abilitata
Backport	HIVE-23617 : corretti i problemi relativi a storage-api FindBug
Backport	HIVE-26408 : vettorizzazione: correzione della deallocazione delle colonne scratch, nessun riutilizzo di una <code>ConstantVectorExpression</code> secondaria come output
Backport	HIVE-23614 : passare sempre HiveConfig a <code>removeTempOrDuplicateFiles</code>
Backport	HIVE-23354 : rimozione del controllo dell'integrità delle dimensioni dei file da <code>CompareTempOrDuplicateFiles</code>

Tipo	Descrizione
Backport	HIVE-20344 : corretto PrivilegeSynchronizer per SBA che genera AccessControlException. È stata inoltre introdotta la proprietà hive.privilege.synchronizer per disabilitare la sincronizzazione dei privilegi
Backport	HIVE-15826 : supporto della configurazione di "serialization.encoding" per tutti i SerDes
Backport	HIVE-18284 : correzione dell'NPE durante l'inserimento di dati con la clausola "distributed by" con ottimizzazione dell'ordinamento in dynpart
Backport	HIVE-24930 : il cortocircuito Operator.setDone() da child op non viene utilizzato nel codepath vettorizzato (se childSize == 1)
Backport	HIVE-24523 : il percorso di lettura vettorizzato per LazySimpleSerde non rispetta SERDEPROPERTIES per il timestamp
Backport	HIVE-23265 : i set di righe duplicati vengono restituiti con Limit e Offset impostati
Backport	HIVE-21492 : VectorizedParquetRecordReader non è in grado di leggere il file di parquet generato utilizzando lo strumento thrift/custom
Backport	HIVE-22540 : vettorizzazione: le colonne Decimal64 non funzionano con VectorizedBatchUtil.makeLikeColumnVector()
Backport	HIVE-22588 : svuota le righe rimanenti per il resto dei set di raggruppamento quando si cambia la modalità del vettore groupby

Tipo	Descrizione
Backport	HIVE-22551 : BytesColumnVector initBuffer dovrebbe pulire il vettore e la lunghezza in modo coerente
Backport	HIVE-22448 : CBO: espandi il conteggio multiplo distinto con una chiave group-by
Backport	HIVE-22248 : correzione dei problemi persistenti delle statistiche
Backport	HIVE-22210 : la vettorizzazione può riutilizzare le colonne di output di calcolo coinvolte nel filtraggio
Backport	HIVE-21531 : vettorizzazione: tutti gli hashcode NULL non vengono calcolati utilizzando Murmur3
Backport	HIVE-20419 : vettorizzazione: impedisce la mutazione di VectorPartitionDesc dopo essere stato utilizzato in una chiave hashmap
Backport	HIVE-19388 : ClassCastException durante l'inizializzazione di VectorMapJoinCommo nOperator
Backport	HIVE-21584 : preparazione a Java 11: il loader di classi di sistema non è URLClassLoader
Backport	HIVE-25107 : la registrazione di Classpath deve essere al livello DEBUG (#2271)
Backport	HIVE-22097 : java.util.ArrayList incompatibile per java 11

Tipo	Descrizione
Backport	HIVE-23938 : LLAP: JDK11 - alcuni argomenti jvm relativi alla rotazione dei file di log GC non possono più essere utilizzati
Backport	HIVE-26226 : esclusione di jdk.tools dep da hive-metastore in upgrade-acid
Backport	HIVE-17879 : aggiornamento del plugin Datanucleus Maven
Backport	HIVE-27004 : DateTimeFormatterBuilder#appendZoneText non può analizzare "UTC+" nelle versioni Java successive a 8
Backport	HIVE-16812 : VectorizedOrcAcidRowBatchReader non filtra gli eventi di eliminazione
Backport	HIVE-17917 : ottimizzazione di VectorizedOrcAcidRowBatchReader.computeOffsetAndBucket
Backport	HIVE-19985 : ACID: saltare la decodifica delle sezioni ROW__ID per le query di sola lettura
Backport	HIVE-20635 : VectorizedOrcAcidRowBatchReader non filtra gli eventi di eliminazione per i file originali
Upgrade	Aggiornamento di Javadoc a 3.3.1
Upgrade	Aggiornamento di Javassist a 3.24.1-GA
Upgrade	Aggiornamento di apache-directory-server a 2.0.0-M14

Nuove configurazioni

Nome	Classificazione	Descrizione
hive.metastore.fs.drop.partition.threads	hive-site	Numero di thread principali nel pool di thread della partizione eliminata.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Tempo in secondi in cui un thread asincrono della partizione inattiva (dal pool di thread) attende l'arrivo di una nuova attività prima di terminare.
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Dimensione massima della coda da utilizzare nel pool di thread per eliminare le partizioni dal file system.
hive.groupby.enable.deterministic.distribution	hive-site	Abilita la distribuzione deterministica delle chiavi sui riduttori. Passa un valore seed costante durante la chiamata alla funzione rand utilizzata per il partizionamento casuale.
hive.privilege.synchronizer	hive-site	Se sincronizzare periodicamente i privilegi dall'auto rizzatore esterno in HiveServer2.
hive.cli.query.file.encoding	hive-site	Codifica dei file per tutti i tipi di file di query (file di query, file di query init, file rc ecc.) forniti negli argomenti CLI.

Nome	Classificazione	Descrizione
hive.emr.tez.shuffle.enabled	hive-site	I lavori Hive su Tez ora utilizzano <code>tez_shuffle</code> per impostazione predefinita invece di <code>mapreduce_shuffle</code> come Shuffle Handler predefinito.

Configurazioni obsolete

Le seguenti proprietà di configurazione sono obsolete a causa di [HIVE-23354](#) e non sono più supportate con Amazon EMR 6.11.0 e versioni successive.

Nome	Valore predefinito
hive.mapred.reduce.tasks speculative.execution	false
tez.am.speculation.enabled	false

Amazon EMR 6.10.0: note di rilascio per Hive

Amazon EMR 6.10.0: modifiche di Hive

Tipo	Descrizione
Caratteristica	Abilitazione dei controlli di accesso basati su AWS Lake Formation per le query Apache Hive (scrittura) tramite IAM Passthrough (API HiveCLI/Steps) .
Miglioramenti	Disabilitazione della configurazione <code>hive.log.explain.output</code> per impostazione predefinita per ridurre le dimensioni del log

Tipo	Descrizione
Backport	HIVE-26408 : vettorizzazione: correzione della deallocazione delle colonne scratch, nessun riutilizzo di una ConstantVectorExpression secondaria come output
Backport	HIVE-22269 : correzione del conteggio errato dei riduttori nelle query di inserimento con partizione dinamica a causa della mancanza di statistiche provocata da HIVE-20703 .
Backport	HIVE-22891 : omissione del passaggio di estrazione PartitionDesc in CombineHiveRecord per la modalità di esecuzione non LLAP
Backport	HIVE-23804 : aggiunta del database predefinito per le tabelle specifiche delle statistiche delle colonne nello schema del metastore Hive per renderle compatibili con le versioni precedenti
Backport	HIVE-25277 : cancellazione lenta delle partizioni Hive per gli archivi di oggetti cloud con ListFiles costosi
Backport	HIVE-19202 : CBO con esito negativo a causa di NullPointerException in HiveAggregate.isBucketedInput()
Backport	HIVE-19048 : gli errori di correzione di Initscript beeline vengono ignorati
Backport	HIVE-21085 : il registro delle viste materializzate avvia una sessione di tez non esterna

Tipo	Descrizione
Backport	HIVE-21675 : CREATE VIEW IF NOT EXISTS restituisce un errore anziché "OK" se la vista esiste già. Questa è una regressione rispetto a Hive 2.
Backport	HIVE-21646 : Tez: impossibilità di TezTasks di evitare il contesto di registrazione dei thread
Backport	HIVE-22054 : possibilità di evitare l'elenco ricorsivo per verificare se una directory è vuota
Backport	HIVE-16587 : NPE quando si inseriscono tipi complessi con valori nulli annidati
Backport	HIVE-22647 : abilitazione del pool di sessioni per impostazione predefinita
Backport	HIVE-13288 : messaggio di eccezione ambiguo in DagUtils.localizeResource
Backport	HIVE-23870 : ottimizza più conversioni di testo in WritableHiveCharObjectInspector.getPimitiveJavaObject / HiveCharWritable
Backport	HIVE-21498 : aggiornamento di Thrift a 0.13.0
Backport	HIVE-24378 : gli spazi iniziali e finali non vengono rimossi prima della conversione decimale
Backport	HIVE-21341 : valori predefiniti: hive.server2.idle.operation.timeout e hive.server2.idle.session.timeout sono troppo alti
Backport	HIVE-22465 : aggiunta di ssl conf in TezConfigurationFactory

Tipo	Descrizione
Backport	HIVE-24710 : ottimizzazione dell'iterazione PTF per count(*) per ridurre i costi di CPU e I/O
Backport	HIVE-15406 : considerazione della possibilità di vettorializzare la nuova funzione "trunc"
Backport	HIVE-21541 : correzione delle intestazioni asf mancanti da HIVE-15406
Backport	HIVE-24808 : salvataggio nella cache delle date analizzate
Backport	HIVE-24746 : PTF: TimestampValueBoundaryScanner può essere ottimizzato durante il calcolo dell'intervallo
Backport	HIVE-25059 : l'evento alter viene convertito per essere rinominato durante la replica
Backport	HIVE-25142 : rehashing in map join nella tabella di hash rapido che causa il danneggiamento di chiavi di grandi dimensioni
Backport	HIVE-23756 : aggiunti altri vincoli al file package.jdo
Backport	HIVE-25150 : i caratteri di tabulazione non vengono rimossi prima della conversione decimale, analogamente al carattere di spazio che è stato corretto come parte di HIVE-24378
Backport	HIVE-25093 : date_format() UDF restituisce l'output solo nel fuso orario UTC
Backport	HIVE-25268 : date_format udf restituisce risultati errati per le date precedenti al 1900 se il fuso orario locale è diverso da UTC

Tipo	Descrizione
Backport	HIVE-25338 : AIOBE in conv UDF se l'input è vuoto
Backport	HIVE-22400 : il minuto UDF con il tempo restituisce NULL
Backport	HIVE-25058 : PTF: TimestampValueBoundaryScanner può essere ottimizzato durante il calcolo dell'intervallo pt2 - isDistanceGreater
Backport	HIVE-25449 : datediff() fornisce un output errato quando viene eseguito in un'attività tez con un fuso orario diverso da UTC
Backport	HIVE-23688 : vettorizzazione: IndexArrayOutOfBoundsException per una colonna di tipo mappa che include un valore nullo
Backport	HIVE-22247 : HiveHFileOutputFormat genera FileNotFoundException quando l'output dell'attività della partizione è vuoto
Backport	HIVE-25570 : Hive deve inviare il percorso URL completo per l'autorizzazione per la posizione del comando insert overwrite
Backport	HIVE-22903 : row_number() vettorizzato reimposta il numero di riga dopo un batch in caso di espressione costante nella clausola di partizione
Backport	HIVE-25549 : risultati errati per la funzione finestra con espressione nella clausola PARTITION BY o ORDER BY
Backport	HIVE-25579 : la sovrascrittura LOAD aggiunge anziché sovrascrivere

Tipo	Descrizione
Backport	HIVE-25659 : le query sql dirette di Metastore con IN/(NOT IN) devono essere suddivise in base ai parametri massimi consentiti da SQL DB
Backport	HIVE-20502 : correzione di NPE durante l'esecuzione di skewjoin_mapjoin10.q quando vengono utilizzate le statistiche delle colonne.
Backport	HIVE-25765 : la proprietà skip.header.line.count salta le righe di ogni blocco in FetchOperator quando le dimensioni del file sono maggiori
Bug	Correzione di NPE all'inserimento in determinati scenari quando <code>hive.stats.column.autogather</code> e <code>hive.groupby.skewindata</code> sono entrambi abilitati
Bug	Correzione di NPE quando il valore <code>mapred.tasktracker.expiry.interval</code> non è impostato

Amazon EMR 6.9.0 - Note sul rilascio per Hive

Amazon EMR 6.9.0 - Modifiche a Hive

Tipo	Descrizione
Upgrade	Aggiornamento di Jetty a 9.4.48.v20220622
Upgrade	Supporto per Hadoop 3.3.3
Caratteristica	Integrazione di Amazon EMR Hive con Lake Formation per carichi di lavoro interattivi utilizzando l'API di GCSC.

Tipo	Descrizione
Caratteristica	Integrazione di Amazon EMR Hive con Iceberg.
Miglioramenti	Abilita SSL in HiveServer2 quando la crittografia in transito è abilitata utilizzando le configurazioni di sicurezza di Amazon EMR.
Miglioramenti	Abilita il committer ottimizzato per Amazon S3 di Hive EMRFS per impostazione predefinita. Per ulteriori informazioni, consulta Abilitazione del committer ottimizzato EMRFS S3 di Hive .
Miglioramenti	Aggiungi HiveHBaseTableInputFormatV2, che eredita solo la versione mapred di InputFormat per corregger e SPARK-34210 . Imposta <code>hive.hbase.inputformat.v2</code> su <code>true</code> per utilizzarlo.
Miglioramenti	Attendi che TezAM si avvii in background con hive.cli.tez.session.async invece di terminare la sessione e avviarne immediatamente un'altra. Utilizza <code>hive.emr.cli.tez.session.open.timeout</code> per impostare questo timeout in secondi.
Miglioramenti	Aggiungi l'opzione hive.conf.restricted.list.append per aggiungere configurazioni separate da virgole all'elenco di configurazioni con restrizioni <code>hive.conf.restricted.list</code> esistente.
Miglioramenti	È stato reso più chiaro il messaggio di errore restituito quando la query Hive fallisce perché la posizione non è definita per il database.

Tipo	Descrizione
Backport	HIVE-24484 : aggiornamento di Hadoop a 3.3.1 e Tez a 0.10.2
Backport	HIVE-22398 : è stata rimossa la gestione delle code YARN tramite ShimLoader.
Backport	HIVE-23190 : (LLAP) è stato modificato IndexCache per passare l'oggetto del file system a TezSpillRecord.
Backport	HIVE-22185 : HADOOP-15832 causerà problemi con i test che utilizzano i cluster MiniYarn.
Backport	HIVE-21670 : mockito-all è stata sostituita con la dipendenza mockito-core .
Backport	HIVE-24542 : Guava è stato preparato per gli aggiornamenti.
Backport	HIVE-23751 : (JUnit) il metodo #mkdirs() è stato sovrascritto in ProxyFileSystem per l'allineamento in seguito a HADOOP-16582.
Backport	HIVE-21603 : preparazione di Java 11: aggiornamento della versione powermock.
Backport	HIVE-24083 : errore hcatalog, è necessario il tipo di autenticazione, in Hadoop 3.3.0.
Backport	HIVE-24282 : le colonne di visualizzazione non devono ordinare le colonne di output a meno che non sia esplicitamente indicato.
Backport	HIVE-20656 : (impostazioni predefinite sensibili) le configurazioni della memoria di aggregazione delle mappe sono troppo aggressive.

Tipo	Descrizione
Backport	HIVE-25443 : Arrow SerDe non è in grado di serializzare/deserializzare tipi di dati complessi quando sono presenti più di 1.024 valori.
Backport	HIVE-19792 : orc è stato aggiornato al rilascio 1.5.2 e sono stati abilitati i test di evoluzione dello schema decimal_64.
Backport	HIVE-20437 : l'evoluzione dello schema può essere gestita da float, double e decimal.
Backport	HIVE-21987 : Hive non è in grado di leggere Parquet int32 annotato con decimali.
Backport	HIVE-20038 : l'aggiornamento delle query su tabelle non inserite in bucket e partizionate genera NPE.

Amazon EMR 6.9.0 - Problemi noti di Hive

- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su -1 per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.

Amazon EMR 6.8.0 - Note sul rilascio di Hive

Amazon EMR 6.8.0 - Modifiche a Hive

Tipo	Descrizione
Miglioramenti	Riduci le chiamate al file system nel comando msck. Miglioramenti delle prestazioni (~15-20 volte su oltre 10.000 partizioni)
Backport	HIVE-20678 : HiveHBaseTableOutputFormat dovrebbe implementare HiveOutputFormat per garantire la compatibilità
Backport	HIVE-21040 : msck elenca i file non necessari all'ultimo livello dell'albero di directory
Backport	HIVE-21460 : il caricamento dei dati seguito da una query seleziona * produce risultati errati
Backport	HIVE-21660 : risultato errato quando si utilizza l'unione completa e la vista successiva con explode
Backport	HIVE-22505 : ClassCastException causata da una selezione errata dell'operatore vettorizzato
Backport	HIVE-22513 : la propagazione costante della colonna generata nelle operazioni di filtro può causare risultati errati
Backport	HIVE-23435 : mancano le righe al risultato completo dell'outer join
Backport	HIVE-24209 : conversione errata degli argomenti di ricerca per l'operazione NOT BETWEEN quando la vettorizzazione è abilitata

Tipo	Descrizione
Backport	HIVE-24934 : l'annotazione VectorizedExpressions non è necessaria in GenericUDFSQCountCheck
Backport	HIVE-25278 : HiveProjectJoinTransposeRule può eseguire trasformazioni non valide con espressioni di finestra
Backport	HIVE-25505 : risultati errati con l'intestazione skip.header.line.count se la prima riga è vuota
Backport	HIVE-26080 : Aggiorna accumulo-core a 1.10.1
Backport	HIVE-26235 : La condizione OR sulla colonna binaria restituisce un risultato vuoto
Bug	Corregge più registri di avvisi relativi ai binding SLF4J in stderr durante l'avvio
Bug	Corregge l'errore della query SHOW TABLE EXTENDED con l'errore FS errato quando la partizione e la tabella si trovano su file.

Amazon EMR 6.8.0 - Problemi noti di Hive

- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.

Amazon EMR 6.7.0 - Note sul rilascio di Hive

Amazon EMR 6.7.0 - Modifiche a Hive

Tipo	Descrizione
Caratteristica	Integrazione di Amazon EMR con LakeFormation.
Caratteristica	Registrazione di verifica aggiuntiva per il committer ottimizzato per Amazon S3 di Hive EMRFS. Configurazione Hive: <code>hive.blobstore.output-committer.logging</code> , predefinita: <code>false</code>
Caratteristica	Directory di destinazione eliminata durante la sovrascrittura dell'inserimento con un risultato di selezione vuoto in una tabella non partizionata/partizione statica per comportarsi in modo simile a Hive 2.x. Configurazione Hive: <code>hive.emr.iow.clean.target.dir</code> , predefinita: <code>false</code>
Bug	Risolto un errore di query intermittente quando si utilizzava il committer ottimizzato per Amazon S3 di Hive EMRFS con ordinamento del bucket di partizione.
Upgrade	Hive aggiornato alla versione 3.1.3. Fai riferimento a Note di rilascio di Apache Hive 3.1.3 per maggiori dettagli.
Upgrade	Parquet aggiornato a 1.12.2 .
Backport	HIVE-20065 : il metastore non dovrebbe basarsi su jackson 1.x
Backport	HIVE-20071 : esegui la migrazione a jackson 2.x e impedisce l'utilizzo

Tipo	Descrizione
Backport	HIVE-20607 : TXNHandler dovrebbe utilizzare PreparedStatement per eseguire query SQL dirette
Backport	HIVE-20740 : rimuovi il blocco globale nel metodo ObjectStore.setConf
Backport	HIVE-20961 : ritiro dell'implementazione NVL
Backport	HIVE-22059 : il file jar hive-exec non contiene la libreria jackson (fasterxml)
Backport	HIVE-22351 : correggi l'utilizzo errato di ObjectStore con thread in TestObjectStore
Backport	HIVE-23534 : NPE in RetryingMetaStoreClient#invoke durante la cattura di MetaException senza messaggio
Backport	HIVE-24048 : armonizza i componenti Jackson alla versione 2.10.latest - Hive
Backport	HIVE-24768 : utilizza ovunque jackson-bom per la sostituzione della versione
Backport	HIVE-24816 : aggiorna jackson a 2.10.5.1 o 2.11.0+ a causa di CVE-2020-25649
Backport	HIVE-25971 : l'arresto dell'attività Tez viene ritardato a causa del pool di thread memorizzato nella cache non chiuso
Backport	HIVE-26036 : NPE causato da getMTable () in ObjectStore

Amazon EMR 6.7.0 - Problemi noti di Hive

- Le query con funzioni finestra sulla stessa colonna di join possono portare a trasformazioni non valide, come riportato in [HIVE-25278](#) e causare risultati errati o errori di query. Una soluzione alternativa sarebbe disabilitare CBO a livello di query per tali query. La correzione sarà disponibile in una versione di Amazon EMR successiva alla 6.7.0. Per ulteriori informazioni, contatta il Supporto AWS.
- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su `-1` per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.

Amazon EMR 6.6.0 - Note sul rilascio di Hive

Amazon EMR 6.6.0 - Modifiche a Hive

Tipo	Descrizione
Upgrade	Aggiorna Parquet a 1.12.1 .
Upgrade	Aggiorna la versione dei jar jetty a 9.4.43.v20210629
Bug	È stato risolto un problema che causava l'installazione di Hive su tutti i nodi task/principali quando LLAP era abilitato in un cluster Hive.
Backport	HIVE-25942 : aggiornare commons-io alla versione 2.8.0 per via di CVE-2021-29425
Backport	HIVE-25726 : aggiornamento della velocità a 2.3 grazie a CVE-2020-13936

Tipo	Descrizione
Backport	HIVE-25680 : autorizzare l'API #get_table_meta HiveMetastore Server per consentire di utilizzare uno qualsiasi dei modelli di autorizzazione di HiveMetastore.
Backport	HIVE-25554 : aggiornare versione arrow alla 0.15
Backport	HIVE-25242 : la query funziona con estrema lentezza con vectorized.adaptor selezionato
Backport	HIVE-25085 : i client MetaStore non sono più condivisi tra le sessioni.
Backport	HIVE-24827 : la query di aggregazione Hive restituisce risultati non corretti per i file non di testo.
Backport	HIVE-24683 : Hadoop23Shim getFileId incline a NPE per percorsi inesistenti
Backport	HIVE-24656 : CBO ha esito negativo per le query con "is null" ("è nullo") sui tipi di mappa e array
Backport	HIVE-24556 : ottimizzare DefaultGraphWalker per il caso senza nipote
Backport	HIVE-24408 : aggiornare Parquet alla versione 1.11.1
Backport	HIVE-24391 : risoluzione errori file TestOrcFile in branch-3.1
Backport	HIVE-24362 : l'elaborazione dell'albero AST non è ottimale per l'albero con un numero elevato di nodi

Tipo	Descrizione
Backport	HIVE-24316 : aggiornamento ORC da 1.5.6 a 1.5.8 in branch-3.1
Backport	HIVE-24307 : Beeline con file di proprietà e parametro -e ha esito negativo
Backport	HIVE-24245 : PTF vettorizzato con conteggio e distinto su partizione che produce risultati errati.
Backport	HIVE-24224 : correggere il salto di intestazione/ piè di pagina per Hive su Tez su file compressi
Backport	HIVE-24157 : modalità strict non riuscirà su CAST timestamp ↔ numerico
Backport	HIVE-24113 : NPE nel GenericUDFToUnixTimeStamp
Backport	HIVE-23987 : aggiornare versione arrow alla 0.11.0
Backport	HIVE-23972 : aggiungere ID client esterno al client esterno LLAP
Backport	HIVE-23806 : evitare di cancellare gli stati delle statistiche delle colonne in tutte le partizioni nel caso in cui lo schema sia esteso. Ciò migliora il runtime dell'istruzione "alter table add columns" ("modifica tabelle aggiungi colonne").
Backport	HIVE-23779 : BasicStatsTask Info non stampato nella console di beeline
Backport	HIVE-23306 : il comando RESET non funziona se è presente una configurazione impostata da System.getProperty

Tipo	Descrizione
Backport	HIVE-23164 : il server non è terminato correttamente a causa di thread non daemon
Backport	HIVE-22967 : supportare hive.reloadable.aux.jars.path per Hive su Tez
Backport	HIVE-22934 : contatori di log interattivi del server Hive per il flusso di errori
Backport	HIVE-22901 : la sostituzione delle variabili può portare a OOM sui riferimenti circolari
Backport	HIVE-22769 : risultati delle query errati e query non riuscite durante la suddivisione dei file di testo compressi
Backport	HIVE-22716 : lettura su ByteBuffer interrotta in ParquetFooterInputFromCache
Backport	HIVE-22648 : aggiornare Parquet alla versione 1.11.0
Backport	HIVE-22640 : Decimal64columnVector: ClassCastException quando il tipo di colonna di partizione è decimale
Backport	HIVE-22621 : test case instabile: TestLlapSignerImpl.TestSigning
Backport	HIVE-22533 : correggere le possibili vulnerabilità dell'interfaccia utente web del daemon LLAP
Backport	HIVE-22532 : PTFPPD può spingere il limite in modo errato tramite la funzione Rank/DenseRank

Tipo	Descrizione
Backport	HIVE-22514 : HivePrototologgingHook potrebbe consumare molta memoria
Backport	HIVE-22476 : la funzione Hive datediff ha fornito risultati incoerenti quando hive.fetch.task.conversion è impostato su "none" ("nessuno")
Backport	HIVE-22429 : le tabelle in cluster migrate utilizzando bucketing_version 1 su hive 3 utilizza bucketing_version 2 per gli inserti
Backport	HIVE-22412 : StatsUtils lancia NPE durante la descrizione
Backport	HIVE-22360 : MultiDelimitSerDe restituisce risultati errati nell'ultima colonna quando il file caricato ha più colonne di quelle nello schema di tabella
Backport	HIVE-22332 : Hive dovrebbe garantire impostazioni valide per l'evoluzione dello schema a partire da ORC-540
Backport	HIVE-22331 : unix_timestamp senza argomento restituisce timestamp in millisecondi anziché in secondi
Backport	HIVE-22275 : OperationManager.queryIdOperation non pulisce correttamente più queryID
Backport	HIVE-22273 : il controllo di accesso non è riuscito quando è stata rimossa una directory temporanea
Backport	HIVE-22270 : aggiornare commons-io alla versione 2.6

Tipo	Descrizione
Backport	HIVE-22241 : implementare UDF per interpretare data/ora utilizzando la sua rappresentazione interna e il calendario ibrido gregoriano-giuliano
Backport	HIVE-22241 : implementare UDF per interpretare data/ora utilizzando la sua rappresentazione interna e il calendario ibrido gregoriano-giuliano
Backport	HIVE-22232 : NPE quando hive.order.columnalignment è impostato su false
Backport	HIVE-22231 : la query hive con grandi dimensioni tramite Knox ha esito negativo con "Broken pipe Write" non riuscito
Backport	HIVE-22221 : client esterno llap: è necessario ridurre LlapBaseInputFormat#getSplits
Backport	HIVE-22208 : il nome della colonna con parola chiave riservata non viene salvato quando la query che include il join su una tabella con la colonna mascherata viene riscritta
Backport	HIVE-22197 : Common Merge join genera un'eccezione cast di classe.
Backport	HIVE-22170 : from_unixtime e unix_time stamp dovrebbero utilizzare il fuso orario della sessione utente
Backport	HIVE-22169 : Tez: SplitGenerator tenta di cercare file di piano che non esisteranno per Tez

Tipo	Descrizione
Backport	HIVE-22168 : rimuovere la registrazione molto onerosa dall'hotpath della cache llap
Backport	HIVE-22161 : UDF: FunctionRegistry si sincronizza sulla classe org.apache.hadoop.hive ql.udf.UDFType
Backport	HIVE-22120 : correggere l'eccezione results/ArrayOutOfBoundsException errati nei join mappa esterna rimasti in condizioni di confine specifiche
Backport	HIVE-22115 : impedire la creazione di appender di routing di query se la proprietà è impostata su false
Backport	HIVE-22113 : impedire l'arresto di LLAP su RuntimeException correlato a AMReporter
Backport	HIVE-22106 : rimuovere la sincronizzazione tra query per la valutazione della partizione
Backport	HIVE-22099 : varie FDU correlate alla data non possono gestire correttamente le date del calendario giuliano dall'HIVE-20007
Backport	HIVE-22037 : HS2 dovrebbe registrare quando si disattiva a causa di OOM
Backport	HIVE-21976 : l'offset dovrebbe essere nullo anziché zero in Calcite HiveSortLimit
Backport	HIVE-21924 : dividere i file di testo anche se esiste intestazione/piè di pagina
Backport	HIVE-21913 : GenericUDTFGetSplits dovrebbe gestire i nomi utente allo stesso modo di LLAP

Tipo	Descrizione
Backport	HIVE-21905 : miglioramento dei parametri generici rispetto alla classe FetchOperator
Backport	HIVE-21902 : interfaccia utente HiveServer2: l'intestazione della risposta jetty necessita di X-Frame-Options
Backport	HIVE-21888 : impostare hive.parquet.timesamp.skip.conversion predefinito su true
Backport	HIVE-21868 : vettorizzazione CAST... FORMAT
Backport	HIVE-21864 : LlapBaseInputFormat#closeAll
Backport	HIVE-21863 : migliorare il casting del tipo vettorizzato per l'espressione WHEN
Backport	HIVE-21862 : ORC ppd produce risultati errati con timestamp
Backport	HIVE-21846 : creare un thread in TezAM che recupera periodicamente le metriche di LlapDaemon
Backport	HIVE-21837 : MapJoin genera un'eccezione quando la colonna selezionata ha valori completamente nulli
Backport	HIVE-21834 : evitare chiamate inutili per semplificare le condizioni del filtro
Backport	HIVE-21832 : nuove metriche per ottenere il tempo medio di queue/serving/response
Backport	HIVE-21827 : più chiamate in SemanticAnalyzer non passano attraverso il metodo getObjectByName

Tipo	Descrizione
Backport	HIVE-21822 : esporre le metriche LlapDaemon tramite un nuovo metodo API
Backport	HIVE-21818 : CBO: la copia di TableReloptHiveTable contiene traffico di metastore
Backport	HIVE-21815 : le statistiche nel file ORC vengono analizzate due volte
Backport	HIVE-21805 : HiveServer2: utilizzare le API di ShutdownHookManager veloci
Backport	HIVE-21799 : NullPointerException in DynamicPartitionPruningOptimization quando la chiave di join è nella colonna di aggregazione
Backport	HIVE-21794 : aggiungere parametri di vista materializzata a sqlStdAuthSafeVarNameRegexes
Backport	HIVE-21768 : JDBC: eliminare il prefisso dell'unione predefinito per le query UNION non incluse
Backport	HIVE-21746 : ArrayIndexOutOfBoundsException durante l'hash join partizionato dinamicamente, con CBO disabilitato
Backport	HIVE-21717 : la rinominazione non è riuscita per la directory nell'attività di spostamento.
Backport	HIVE-21685 : semplificazione errata nella query con più clausole IN
Backport	HIVE-21681 : la descrizione formattata mostra informazioni errate per più chiavi primarie

Tipo	Descrizione
Backport	HIVE-21651 : spostare protobuf serde in hive-exec.
Backport	HIVE-21619 : tipo di stampa timestamp impreciso nella spiegazione SQL estesa
Backport	HIVE-21592 : OptimizedSql non viene visualizzato se l'espressione contiene CONCAT
Backport	HIVE-21576 : presentazione di CAST... FORMAT e di un elenco limitato di formati data/ora SQL:2016
Backport	HIVE-21573 : il trasporto binario ignora il principale se l'autenticazione è impostata su delegationToken
Backport	HIVE-21550 : test TestObjectStore errati: impossibile ottenere un blocco entro il tempo richiesto
Backport	HIVE-21544 : la propagazione costante corrompe le espressioni coalesce/case/when durante il fold
Backport	HIVE-21539 : GroupBy + dove la clausola sulla stessa colonna provoca una riscrittura errata della query
Backport	HIVE-21538 : Beeline: origine password attraverso il lettore della console non passata al parametro di connessione
Backport	HIVE-21509 : LLAP potrebbe memorizzare nella cache i vettori di colonna danneggiati e restituire un risultato errato

Tipo	Descrizione
Backport	HIVE-21499 : non deve rimuovere la funzione dal registro se il comando "create" ("crea") non è riuscito con AlreadyExistsException
Backport	HIVE-21496 : possibile overflow del buffer non ordinato durante il dimensionamento automatico
Backport	HIVE-21468 : distinzione tra maiuscole e minuscole nei nomi degli identificatori per il gestore di archiviazione JDBC
Backport	HIVE-21467 : rimozione importazioni junit.framework.Assert obsolete
Backport	HIVE-21435 : LlapBaselInputFormat dovrebbe ottenere il numero di attività da TASK_ATEMPT_ID conf, se presente, durante la creazione di SubmitWorkRequestProto
Backport	HIVE-21389 : la distribuzione di Hive non presenta javax.ws.rs-api.jar dopo HIVE-21247
Backport	HIVE-21385 : autorizzazione a disabilitare il pushdown del calcolo non divisibile su sorgenti JDBC
Backport	HIVE-21383 : gestore di archiviazione JDBC: utilizzare catalogo e schema per recuperare le tabelle se specificato
Backport	HIVE-21382 : ottimizzazione della riduzione dei gruppi per chiavi: chiavi non ridotte su query23
Backport	HIVE-21362 : aggiungere un formato di input e serde per leggere dai file protobuf.

Tipo	Descrizione
Backport	HIVE-21340 : CBO: eliminazione colonne non chiave che alimentano una SemiJoin
Backport	HIVE-21332 : eliminazione buffer non bloccati anziché buffer bloccati
Backport	HIVE-21329 : dimensioni del buffer di output non ordinato del runtime Tez personalizzate a seconda della pipeline dell'operatore
Backport	HIVE-21295 : StorageHandler converte la data in stringa utilizzando la convenzione Hive
Backport	HIVE-21294 : vettorizzazione: lo shuffle con 1 riduttore può saltare le funzioni hash dell'oggetto
Backport	HIVE-21255 : rimuovere QueryConditionBuilder in JdbcStorageHandler
Backport	HIVE-21253 : supportare DB2 in JDBC StorageHandler
Backport	HIVE-21232 : LLAP: aggiungere un provider di affinità split friendly per la mancata cache
Backport	HIVE-21214 : MoveTask: usare attemptId anziché le dimensioni del file per la deduplica dei file compareTempOrDuplicateFiles
Backport	HIVE-21184 : aggiunta spiegazione e spiegazione piano CBO formattato con informazioni sui costi
Backport	HIVE-21182 : saltare la configurazione di Hive Scratch Dir durante la pianificazione

Tipo	Descrizione
Backport	HIVE-21171 : saltare la creazione di scratch per tez se RPC è attivo
Backport	HIVE-21126 : autorizzazione interrogazioni a livello di sessione in LlapBaselInputForma t#getSplit
Backport	HIVE-21107 : errore "Cannot find field" ("Impossibile trovare il campo") durante l'hash join partizionato dinamicamente
Backport	HIVE-21061 : la query CTAS ha esito negativo con IllegalStateException per origine vuota
Backport	HIVE-21041 : NPE, ParseException nell'ottenere lo schema dal piano logico
Backport	HIVE-21013 : JdbcStorageHandler non riesce a trovare la colonna delle partizioni in Oracle
Backport	HIVE-21006 : estendere SharedWorkOptimizer per rimuovere i semijoin quando esiste un'opportunità di riutilizzo
Backport	HIVE-20992 : suddividere la configurazione hive.metastore.dbaccess.ssl.properties in configurazioni più significative
Backport	HIVE-20989 : JDBC - L'API GetOperationStatus + log possono bloccare l'avanzamento della query tramite sleep
Backport	HIVE-20988 : risultati errati per gruppi per query con chiave primaria su più colonne

Tipo	Descrizione
Backport	HIVE-20985 : se gli input operatore seleziona ti sono colonne temporanee, la vettorizzazione può riutilizzarne alcuni come output
Backport	HIVE-20978 : "hive.jdbc.*" deve essere aggiunto a sqlStdAuthSafeVarNameRegexes
Backport	HIVE-20953 : rimozione di una funzione dal relativo registro quando non può essere aggiunta al metastore durante la sua creazione.
Backport	HIVE-20952 : pulizia di VectorizationContext.java
Backport	HIVE-20951 : LLAP: impostare sempre Xms al 50%
Backport	HIVE-20949 : migliorare la stima della cardinalità PKFK nella pianificazione fisica
Backport	HIVE-20944 : non convalidare le statistiche durante la compilazione delle query
Backport	HIVE-20940 : casi bridge in cui la risoluzione del tipo Calcite è più rigorosa di Hive.
Backport	HIVE-20937 : la query jdbc Postgres ha esito negativo con l'errore "LIMIT must not be negative" ("LIMIT non deve essere negativo")
Backport	HIVE-20926 : il suggerimento di riduzione semi join ha esito negativo quando le voci del filtro Bloom sono alte o quando non ci sono statistiche
Backport	HIVE-20920 : utilizzare i vincoli SQL per migliorare l'algoritmo di riordinamento del join

Tipo	Descrizione
Backport	HIVE-20918 : flag per abilitare/disabilitare l'impostazione del calcolo da Calcite nella connessione JDBC
Backport	HIVE-20915 : disponibilità ottimizzazione dinamica partizioni di ordinamento per HoS e MR
Backport	HIVE-20910 : errore di inserimento nella tabella con bucket a causa dell'ottimizzazione dinamica dell'ordinamento partizioni
Backport	HIVE-20899 : l'URI Keytab per il servizio LLAP YARN è restrittivo per supportare solo HDFS
Backport	HIVE-20898 : per le funzioni correlate al tempo, gli argomenti non possono essere trasferiti su un tipo non annullabile
Backport	HIVE-20881 : la propagazione costante semplifica eccessivamente le proiezioni
Backport	HIVE-20880 : aggiornare il valore predefinito per <code>hive.stats.filter.in.min.ratio</code>
Backport	HIVE-20873 : usare l'hash Murmur per <code>VectorHashKeyWrapperTwoLong</code> per ridurre la collisione hash
Backport	HIVE-20868 : SMB Join ha esito negativo in modo intermittente quando <code>TezDummyOperator</code> ha un'operazione figlio in <code>getFinalOp</code> in <code>MapRecordProcessor</code>
Backport	HIVE-20853 : esporre <code>ShuffleHandler.registerDag</code> nell'API del daemon <code>llap</code>

Tipo	Descrizione
Backport	HIVE-20850 : passaggio del case condizionale dalle proiezioni alle tabelle dimensionali, se possibile
Backport	HIVE-20842 : correggere la logica introdotta in HIVE-20660 per stimare le statistiche di raggruppamento in base a una condizione
Backport	HIVE-20839 : errore "Cannot find field" ("Impossibile trovare il campo") durante l'hash join partizionato dinamicamente
Backport	HIVE-20835 : l'interazione tra vincoli e riscrittura VM può creare loop nel pianificatore di Calcite
Backport	HIVE-20834 : utilizzare Hive per le voci QueryResultCache che mantengono il riferimento a SemanticAnalyzer dalla query memorizzata nella cache
Backport	HIVE-20830 : l'asserzione della query dell'intervallo JdbcStorageHandler ha esito negativo in alcuni casi
Backport	HIVE-20829 : la divisione della gamma JdbcStorageHandler genera NPE
Backport	HIVE-20827 : risultati incoerenti per array vuoti
Backport	HIVE-20826 : migliorata la regola HiveSemiJoin per convertire il gruppo join + dal lato sinistro a Left Semi Join
Backport	HIVE-20821 : riscrivere SUM0 nella combinazione SUM+COALESCE

Tipo	Descrizione
Backport	HIVE-20815 : JdbcRecordReader.next non assorbe eccezioni
Backport	HIVE-20813 : udf to_epoch_milli deve supportar e anche il timestamp senza fuso orario.
Backport	HIVE-20804 : ulteriori miglioramenti al raggruppamento per ottimizzazione con vincoli
Backport	HIVE-20792 : l'inserimento di timestamp con zone tronca i dati
Backport	HIVE-20788 : la riduzione estesa dell'SJ potrebbe far tornare indietro le colonne in modo errato durante la creazione
Backport	HIVE-20778 : il riordino delle join potrebbe non essere attivato se tutte le join nel piano sono create dalla logica di decorrelazione
Backport	HIVE-20772 : registrare i contatori CPU per attività in LLAP
Backport	HIVE-20768 : aggiungere UDF della finestra a cascata
Backport	HIVE-20767 : più progetti tra operatori di join possono influire sul riordinamento del join utilizzando i vincoli
Backport	HIVE-20762 : l'intervallo di pulizia NOTIFICATION_LOG è codificato come 60 ed è troppo piccolo

Tipo	Descrizione
Backport	HIVE-20761 : la selezione per l'aggiornamento sulla tabella notification_sequence ha un intervallo di nuovi tentativi e un numero di tentativi troppo piccoli
Backport	HIVE-20751 : aggiornare versione arrow alla 0.10.0
Backport	HIVE-20746 : HiveProtoHookLogger non chiude il file alla fine della giornata.
Backport	HIVE-20744 : utilizzare i vincoli SQL per migliorare l'algoritmo di riordinamento del join
Backport	HIVE-20740 : rimuovere il blocco globale nel metodo ObjectStore.setConf. Questo cherry-pick supporta HIVE-20740 progettato per Hive 3.2 e da 4.x a 3.1.x
Backport	HIVE-20734 : Beeline: quando beeline-site.xml è attivo e la CLI di Hive reindirizza a Beeline, dovrebbe usare il nome utente del sistema/p password fittizia invece di chiederne una
Backport	HIVE-20731 : il file keystore in JdbcStorageHandler dovrebbe essere autorizzato
Backport	HIVE-20720 : aggiungere l'opzione colonna partizione al gestore JDBC
Backport	HIVE-20719 : l'istruzione SELECT fallisce dopo l'UPDATE con l'ottimizzazione e la vettorizzazione di hive.optimize.sort.dynamic.partition attive
Backport	HIVE-20718 : aggiunto il driver perf cli con vincoli

Tipo	Descrizione
Backport	HIVE-20716 : impostare il valore predefinito per <code>hive.cbo.stats.correlated.multi.key.joins</code> su <code>true</code>
Backport	HIVE-20712 : <code>HivePointLookupOptimizer</code> dovrebbe estrarre casi profondi
Backport	HIVE-20710 : il <code>fold</code> costante potrebbe non creare costanti nulle senza tipi
Backport	HIVE-20706 : <code>external_jdbc_table2.q</code> saltuariamente non funzionante
Backport	HIVE-20704 : estendere <code>HivePreFilteringRule</code> per supportare altre funzioni
Backport	HIVE-20703 : mettere l'ottimizzazione della partizione di ordinamento dinamico nella decisione basata sui costi
Backport	HIVE-20702 : tenere conto del sovraccarico derivante dalle stime consapevoli della struttura dei dati durante la selezione di <code>mapjoin</code>
Backport	HIVE-20692 : abilitare il <code>fold</code> delle espressioni <code>NOT x IS (NOT) [TRUE FALSE]</code>
Backport	HIVE-20691 : correggere <code>org.apache.hadoop.hive.cli.TestMiniLlapCliDriver.testCliDriver[cttl]</code>
Backport	HIVE-20682 : l'esecuzione asincrona di query può potenzialmente non riuscire se <code>SessionHive</code> condivisa viene chiusa dal thread principale
Backport	HIVE-20676 : <code>HiveServer2: PrivilegeSynchronizer</code> non è impostato sullo stato del daemon

Tipo	Descrizione
Backport	HIVE-20660 : la stima raggruppa per statistica potrebbe essere migliorata limitando il numero totale di righe alla tabella di origine
Backport	HIVE-20652 : riunificazione push di JdbcStorageHandler per due diverse origini dati al driver jdbc
Backport	HIVE-20651 : la password JdbcStorageHandler deve essere crittografata
Backport	HIVE-20649 : gestore di memoria LLAP per scrittori Orc
Backport	HIVE-20648 : LLAP: Il gruppo vettoriale per operatore dovrebbe utilizzare la memoria per esecutore
Backport	HIVE-20646 : la condizione del filtro di partizion e non viene spinta verso il basso alla query di metastore se il valore è IS NOT NULL
Backport	HIVE-20644 : evitare di esporre informazioni sensibili tramite un'eccezione Hive Runtime
Backport	HIVE-20636 : migliorare la stima del numero di valori nulli dopo il join esterno
Backport	HIVE-20632 : la query con get_splits UDF ha esito negativo se la vista materializzata viene creata sulla tabella interrogata
Backport	HIVE-20627 : le query asincrone concorrenti non funzionano saltuariamente con LockException e causano perdite di memoria

Tipo	Descrizione
Backport	HIVE-20623 : lavoro condiviso: estendere la condivisione delle voci della cache di map-join in LLAP
Backport	HIVE-20619 : inclusione di MultiDelimiterDE in HiveServer2 per impostazione predefinita
Backport	HIVE-20618 : durante la selezione del join, BucketMapJoin potrebbe essere scelto per tabelle senza bucket
Backport	HIVE-20617 : correggere il tipo di costanti nelle espressioni IN per avere il tipo corretto
Backport	HIVE-20612 : creazione di un nuovo flag di correlazione a più chiavi di join per CBO
Backport	HIVE-20603 : errore "Wrong FS" ("FS errato") durante l'inserimento nella partizione dopo aver modificato il file system della posizione della tabella
Backport	HIVE-20601 : EnvironmentContext nullo nell'evento ALTER_PARTITION in DbNotificationListener
Backport	HIVE-20583 : usare il nome host canonico solo per l'autenticazione kerberos in HiveConnection
Backport	HIVE-20582 : rendere configurabile il flush in hive proto logging
Backport	HIVE-20563 : vettorizzazione: l'espressione CASE WHEN ha esito negativo quando il tipo THEN/ELSE e il tipo di risultato sono diversi

Tipo	Descrizione
Backport	HIVE-20558 : cambia il valore predefinito di <code>hive.hashtable.key.count.adjustment</code> a 0,99
Backport	HIVE-20552 : acquisizione dello schema da LogicalPlan più rapida
Backport	HIVE-20550 : cambiare WebHCat per usare Beeline per inviare query Hive
Backport	HIVE-20537 : multi-colonna riunisce le stime con colonne non correlate diverse in CBO e Hive
Backport	HIVE-20524 : il controllo Schema Evolution è interrotto passando dalla versione 2 di Hive alla versione 3 per ALTER TABLE VARCHAR a DECIMAL
Backport	HIVE-20522 : HiveFilterSetOpTransposeRule può generare un errore di asserzione a causa dell'annullabilità dei campi
Backport	HIVE-20521 : HS2 doAs=true ha un problema di autorizzazione con <code>hadoop.tmp.dir</code> , con filesystem MR e S3A
Backport	HIVE-20515 : risultati della query vuoti quando si utilizza la cache dei risultati e la directory temporanea della query, la directory della cache dei risultati in diversi filesystem
Backport	HIVE-20508 : Hive non supporta nomi utente del tipo "user@realm"
Backport	HIVE-20507 : Beeline: aggiungere un comando di utility per recuperare tutti gli uri da <code>beeline-site.xml</code>

Tipo	Descrizione
Backport	HIVE-20505 : aggiornare org.openjdk.jmh:jmh-core alla versione 1.21
Backport	HIVE-20503 : utilizzare stime consapevoli della struttura dati durante la selezione di mapjoin
Backport	HIVE-20498 : supportare il tipo di data per la raccolta automatica delle statistiche delle colonne
Backport	HIVE-20496 : vettorizzazione: IllegalStateException per PTF vettorizzato
Backport	HIVE-20494 : GenericUDFRestrictInformationSchema è interrotto dopo HIVE-19440
Backport	HIVE-20477 : OptimizedSql non viene visualizzato se l'espressione contiene IN
Backport	HIVE-20467 : ammesse le istruzioni IF NOT EXISTS/IF EXISTS nella creazione/drop del piano di risorse
Backport	HIVE-20462 : "CREATE VIEW IF NOT EXISTS" ha esito negativo se la vista esiste già
Backport	HIVE-20455 : log generato da security.authorization.PrivilegeSynchronizer.run
Backport	HIVE-20439 : usare il limite di memoria gonfiato durante la selezione del join per llap
Backport	HIVE-20433 : la conversione implicita da stringa a timestamp è lenta
Backport	HIVE-20432 : riscrivere da BETWEEN a IN per i tipi interi per la stima delle statistiche

Tipo	Descrizione
Backport	HIVE-20423 : impostazione di NULLS LAST come ordinamento nullo predefinito
Backport	HIVE-20418 : LLAP IO potrebbe non gestire file ORC con indice di riga disabilitato correttamente per le query senza colonne selezionate
Backport	HIVE-20412 : NPE in HiveMetaHook
Backport	HIVE-20406 : Nested Coalesce restituisce risultati errati
Backport	HIVE-20399 : CTAS con una posizione di tabella personalizzata che non è completamente qualificata ha esito negativo per le tabelle MM
Backport	HIVE-20393 : riduzione Semijoin: markSemiJoinForDPP si comporta in modo incoerente
Backport	HIVE-20391 : HiveAggregateReduceFunctionRule può dedurre un tipo di ritorno errato durante la decomposizione della funzione di aggregazione
Backport	HIVE-20383 : nome della coda non valido e problemi di sincronizzazione nell'hook degli eventi proto di hive.
Backport	HIVE-20367 : vettorizzazione: supportare lo streaming per PTF AVG, MAX, MIN, SUM
Backport	HIVE-20366 : le stime delle statistiche TPC-DS query78 sono disattivate per il suo filtro nullo
Backport	HIVE-20364 : aggiornare predefinito per hive.map.aggr.hash.min.reduction

Tipo	Descrizione
Backport	HIVE-20352 : vettorizzazione: supportare la funzione di raggruppamento
Backport	HIVE-20347 : hive.optimize.sort.dynamic.partition dovrebbe funzionare con CTAS e VM partizionati
Backport	HIVE-20345 : il database di rilascio potrebbe bloccarsi se le tabelle vengono eliminate da una chiamata diversa
Backport	HIVE-20343 : Hive 3: CTAS non rispetta le transactional_properties
Backport	HIVE-20340 : Druid necessita di CAST espliciti da Timestamp a STRING quando l'output della funzione timestamp viene utilizzato come Strin
Backport	HIVE-20339 : vettorizzazione: sollevare le restrizioni non necessarie che causano la mancata vettorizzazione di un PTF con RANK
Backport	HIVE-20337 : CachedStore: getPartitionsByExpr non sta compilando correttamente l'elenco delle partizioni
Backport	HIVE-20336 : criteri di mascheramento e filtraggio per viste materializzate
Backport	HIVE-20326 : creare vincoli con RELY come predefinito anziché NO RELY
Backport	HIVE-20321 : vettorizzazione: ridurre la dimensione della memoria di 1 col VectorHasKeyWrapper a <1 CacheLine

Tipo	Descrizione
Backport	HIVE-20320 : attivazione flag hive.optimize.remove_sq_count_check
Backport	HIVE-20315 : vettorizzazione: correggere altri problemi NULL/risultati errati ed evitare cast/conversioni inutili
Backport	HIVE-20314 : include la potatura delle partizioni nella riscrittura delle viste materializzate
Backport	HIVE-20312 : consentire ai client arrow di utilizzare il proprio BufferAllocator con LlapOutputFormatService
Backport	HIVE-20302 : LLAP: l'esecuzione non vettorizzata in IO ignora le colonne virtuali, incluso ROW_ID
Backport	HIVE-20300 : VectorFileSinkArrowOperator
Backport	HIVE-20299 : potenziale competizione nel test dell'unità firmatario LLAP
Backport	HIVE-20296 : migliorare HivePointLookupOptimizerRule per consentire l'estrazione da contesti più sofisticati
Backport	HIVE-20294 : vettorizzazione: risoluzione problemi NULL/risultati errati in COALESCE / ELT
Backport	HIVE-20292 : ordine delle join errato in query93 di tpch con il vincolo primario definito
Backport	HIVE-20290 : Lazy inizializza ArrowColumnarBatchSerDe in modo che non allochi buffer durante GetSplits

Tipo	Descrizione
Backport	HIVE-20281 : SharedWorkOptimizer ha esito negativo con l'errore "operator cache contents and actual plan differ" ("differenza tra il contenuto della cache dell'operatore e il piano effettivo")
Backport	HIVE-20277 : vettorizzazione: le espressioni case che restituiscono BOOLEAN non sono supportate per FILTER
Backport	HIVE-20267 : espandere la WebUI per includere il modulo per configurare dinamicamente i livelli di log
Backport	HIVE-20263 : errore di battitura nella variabile HiverEduceExpressionsWithStatsRule
Backport	HIVE-20260 : l'NDV di una colonna non deve essere ridimensionato quando il numero di righe viene modificato dal filtro su un'altra colonna
Backport	HIVE-20252 : riduzione semijoin: i cicli dovuti al ramo semi join possono rimanere inosservati se il lato piccolo della tabella ha una mappa join a monte.
Backport	HIVE-20245 : vettorizzazione: risolvere problemi NULL/risultati errati in BETWEEN/IN
Backport	HIVE-20241 : supporto delle specifiche di partizionamento nelle istruzioni CTAS
Backport	HIVE-20240 : riduzione semijoin: uso della variabile locale per verificare le condizioni della tabella esterna

Tipo	Descrizione
Backport	HIVE-20226 : HMS getNextNotification genera un'eccezione quando la richiesta maxEvents supera max_rows della tabella
Backport	HIVE-20225 : SerDe supporterà il formato binario Teradata
Backport	HIVE-20213 : aggiornamento di Calcite a 1.17.0
Backport	HIVE-20212 : Hiveserver2 in modalità http emette in modo errato la metrica default.General.open_connections
Backport	HIVE-20210 : l'ottimizzatore di recupero semplice dovrebbe condurre a MapReduce quando il filtro sulla colonna non partizione e conversione è minimo
Backport	HIVE-20209 : la connessione al metastore fallisce per il primo tentativo nel dump repl
Backport	HIVE-20207 : vettorizzazione: risolvere problemi NULL/risultati errati in filtra/confronta
Backport	HIVE-20204 : conversione del tipo durante IN
Backport	HIVE-20203 : Arrow SerDe perde un DirectByteBuffer
Backport	HIVE-20197 : vettorizzazione: aggiunto il test DECIMAL_64, aggiunta l'aritmetica di data/intervallo/timestamp e aggiunte altre aggregazioni GROUP BY
Backport	HIVE-20193 : CBOInfo non è presente nel json del piano descrittivo

Tipo	Descrizione
Backport	HIVE-20192 : HS2 con metastore incorporato sta perdendo oggetti JDOPersistenceManager
Backport	HIVE-20183 : l'inserimento da una tabella con bucket può causare la perdita di dati, se la tabella di origine contiene bucket vuoti
Backport	HIVE-20177 : vettorizzazione: riduzione dell'allocazione di KeyWrapper in modalità streaming GroupBy
Backport	HIVE-20174 : vettorizzazione: risoluzione dei problemi NULL / risultati errati nelle funzioni di aggregazione GROUP BY
Backport	HIVE-20172 : StatsUpdater non riuscito con GSS Exception durante il tentativo di connettersi al metastore remoto
Backport	HIVE-20153 : le UDF Count e Sum consumano più memoria in Hive 2+
Backport	HIVE-20152 : reimpostare lo stato db quando il dump repl ha esito negativo, in modo da poter rinominare la tabella
Backport	HIVE-20149 : TestHiveCli ha esito negativo/va in timeout
Backport	HIVE-20130 : migliore registrazione per il sincronizzatore dello schema delle informazioni
Backport	HIVE-20129 : ripristina l'evoluzione dello schema basato sulla posizione per le tabelle orc

Tipo	Descrizione
Backport	HIVE-20118 : SessionStateUserAuthenticator.getGroupNames
Backport	HIVE-20116 : TezTask utilizza il logger padre
Backport	HIVE-20115 : le tabelle Acid non devono utilizzare la scansione del footer per analizzare
Backport	HIVE-20103 : WM: aggregare solo contatori DAG se ne viene utilizzato almeno uno
Backport	HIVE-20101 : BloomKFilter: evitare di utilizzare completamente gli array di byte[] locali
Backport	HIVE-20100 : OpTraits: selezionare OpTraits dovrebbe interrompersi quando viene rilevata una mancata corrispondenza
Backport	HIVE-20098 : statistiche: NPE quando si ottengono le statistiche sulla partizione della colonna Data
Backport	HIVE-20095 : correggere la funzione per l'invio push della computazione alle tabelle esterne jdbc
Backport	HIVE-20093 : LlapOutputFormatService: uso di ArrowBuf con Netty per la contabilità
Backport	HIVE-20090 : estendere la creazione di filtri di riduzione semijoin per scoprire nuove opportunità
Backport	HIVE-20088 : il percorso della posizione di configurazione di Beeline è assemblato in modo errato

Tipo	Descrizione
Backport	HIVE-20082 : la conversione da HiveDecimal a stringa non formatta correttamente il decimale
Backport	HIVE-20069 : correggere la riottimizzazione in caso di ottimizzazione DPP e Semijoin
Backport	HIVE-20051 : ignorare l'autorizzazione per le tabelle temporanee
Backport	HIVE-20044 : Arrow Serde dovrebbe tamponare i valori dei caratteri e gestire correttamente le stringhe vuote
Backport	HIVE-20028 : la configurazione della cache client metastore viene utilizzata in modo errato
Backport	HIVE-20025 : pulizia dei file di eventi creati da HiveProtoLoggingHook
Backport	HIVE-20020 : il jar contrib di Hive non deve essere nella lib
Backport	HIVE-20013 : aggiunta di un cast implicito al tipo di data per la funzione to_date
Backport	HIVE-20011 : allontanarsi dalla modalità di aggiunta nell'hook di registrazione proto
Backport	HIVE-20005 : acid_table_stats, acid_no_buckets, ecc.; modifica del risultato della query sul branch
Backport	HIVE-20004 : la scala errata utilizzata dai risultati ConvertDecimal64ToDecimal produce risultati errati

Tipo	Descrizione
Backport	HIVE-19995 : aggregazione del traffico di righe per le tabelle Acid
Backport	HIVE-19993 : non è possibile utilizzare un alias di tabella che appare anche come nome di colonna
Backport	HIVE-19992 : vettorizzazione: follow-on su HIVE-19951 —> aggiungere la chiamata a SchemaEvolution.isOnlyImplicitConversion per disabilitare l'I/O LLAP codificato per ORC solo quando la conversione del tipo di dati non è implicita
Backport	HIVE-19989 : Metastore utilizza il nome dell'applicazione errato per le metriche HADOOP2
Backport	HIVE-19981 : le tabelle gestite convertite in tabelle esterne dall'utilità HiveStrictManagedMigration devono essere impostate per eliminare i dati quando la tabella viene eliminata
Backport	HIVE-19967 : SMB Join: necessari Optraits per PTFOperator alla GBY Op
Backport	HIVE-19935 : sessione Hive WM eliminata: non è stato possibile aggiornare il conteggio delle attività LLAP
Backport	HIVE-19924 : contrassegno dei processi distcp eseguiti da Repl Load
Backport	HIVE-19891 : l'inserimento in tabelle esterne con directory di partizioni personalizzate può causare la perdita di dati

Tipo	Descrizione
Backport	HIVE-19850 : la potatura dinamica delle partizioni in Tez conduce all'errore "No work found for tablescan" ("Nessun lavoro trovato per la scansione tabelle")
Backport	HIVE-19806 : ordinamento dell'output dei qtest per evitare errori nei risultati dei test
Backport	HIVE-19770 : supporto per CBO per le query con più colonne uguali nella select
Backport	HIVE-19769 : creazione di oggetti dedicati per i nomi di DB e tabelle
Backport	HIVE-19765 : aggiunta di test specifici per Parquet su BlobStoreCliDriver
Backport	HIVE-19759 : test flaky: TestRpc#testServerPort
Backport	HIVE-19711 : rifattorizzare strumento schema di Hive
Backport	HIVE-19701 : getDelegationTokenFromMetaStore non deve essere sincronizzato
Backport	HIVE-19694 : l'istruzione Create Materialized View dovrebbe verificare la presenza di conflitti di nomi MV prima di eseguire l'istruzione SQL di MV.
Backport	HIVE-19674 : raggruppamento per costanti decimali passato poi fino alla tabella Druid
Backport	HIVE-19668 : oltre il 30% dell'heap sprecato da org.antlr.runtime.comMonToken duplicati e da stringhe duplicate

Tipo	Descrizione
Backport	HIVE-19663 : rifattorizzazione della generazione di report LLAP IO
Backport	HIVE-19661 : cambiare le UDF Hive per utilizzar e il motore regex Re2J
Backport	HIVE-19628 : possibile NPE in LLAP testSigning
Backport	HIVE-19568 : HS2 HA attivo/passivo: non consentire la connessione diretta all'istanza HS2 passiva
Backport	HIVE-19564 : vettorizzazione: risolvere problemi NULL/risultati errati negli operatori aritmetici
Backport	HIVE-19552 : abilitare TestMiniDruidKafka CliDriver#druidkafkamini_basic.q
Backport	HIVE-19432 : GetTablesOperation è troppo lento se l'Hive ha troppi database e tabelle
Backport	HIVE-19360 : CBO: aggiunta di un "SQL ottimizzato" all'oggetto QueryPlan
Backport	HIVE-19326 : raccolta automatica delle statistic he: aggregazione errata durante le query UNION
Backport	HIVE-19313 : i test TestJdbcWithDBTokenStoreNoDoAs non funzionano
Backport	HIVE-19285 : aggiunta dei registri alle sottoclas si di MetadataOperation
Backport	HIVE-19235 : aggiornamento dei file golden per i test Minimr

Tipo	Descrizione
Backport	HIVE-19104 : quando il test MetaStore viene avviato con il nuovo tentativo, le istanze devono essere indipendenti
Backport	HIVE-18986 : la ridenominazione della tabella esegue java.lang.StackOverflowError in DataNucleus se la tabella contiene un numero elevato di colonne
Backport	HIVE-18920 : CBO: inizializzare i provider Janino prima della prima query
Backport	HIVE-18873 : saltare il pushdown dei predicati per MR silenziosamente su HiveInputFormat può causare la produzione di risultati errati da parte dei gestori di archiviazione
Backport	HIVE-18871 : errore di esecuzione hive su tez dovuto all'impostazione di hive.aux.jars.path su hdfs://
Backport	HIVE-18725 : miglioramento della gestione degli errori per le query secondarie in caso di riferimento di colonna errato
Backport	HIVE-18696 : le cartelle delle partizioni potrebbero non essere ripulite correttamente nel metodo HiveMetastore.add_Partitions_core se un
Backport	HIVE-18453 : ACID: aggiunta la sintassi "CREATE TRANSACTIONAL TABLE" per unire il supporto di ACID ORC e Parquet
Backport	HIVE-18201 : disabilitata XPROD_EDGE per sq_count_chec

Tipo	Descrizione
Backport	HIVE-18140 : le statistiche delle tabelle partizionate possono risultare errate in un caso combinato di statistiche base
Backport	HIVE-17921 : l'aggregazione con struttura in LLAP produce risultati errati
Backport	HIVE-17896 : TopNKey: creazione di un operatore TopNKey vettorializzabile autonomo
Backport	HIVE-17840 : HiveMetaStore assorbe un'eccezione se transactionalListeners.notifyEvent ha esito negativo
Backport	HIVE-17043 : rimuovere colonne non univoche dal gruppo in base alle chiavi se non viene fatto riferimento in seguito
Backport	HIVE-17040 : partecipare all'eliminazione in presenza di relazione FK
Backport	HIVE-16839 : chiamate sbilanciate a openTransaction/commitTransaction quando si altera contemporaneamente la stessa partizione
Backport	HIVE-16100 : l'ottimizzatore dinamico delle partizioni ordinate perde operatori di pari livello
Backport	HIVE-15956 : StackOverflowError quando si rilasciano molte partizioni
Backport	HIVE-15177 : l'autenticazione con hive ha esito negativo quando il tipo di autenticazione kerberos è impostato su fromSubject e il principale contiene _HOST

Tipo	Descrizione
Backport	HIVE-14898 : HS2 non dovrebbe eseguire il log del callstack per un errore di intestazione di autenticazione vuota
Backport	HIVE-14493 : supporto del partizionamento per le viste materializzate
Backport	HIVE-14431 : riconoscimento di COALESCE come CASE
Backport	HIVE-13457 : creare endpoint HS2 REST API per il monitoraggio delle informazioni
Backport	HIVE-12342 : impostazione del valore predefinito di <code>hive.optimize.index.filter</code> a <code>true</code>
Backport	HIVE-10296 : rilevata un'eccezione di cast quando hive esegue una query multi-join su metastore
Backport	HIVE-6980 : abbandonare tabella usando <code>direct sql</code>

Amazon EMR 6.6.0 - Modifiche alla configurazione di Hive

- Come parte della modifica dell'OSS [HIVE-20703](#), la proprietà di ordinamento delle partizioni dinamiche `hive.optimize.sort.dynamic.partition` è stata sostituita con `hive.optimize.sort.dynamic.partition.threshold`.

La configurazione `hive.optimize.sort.dynamic.partition.threshold` presenta i seguenti valori potenziali:

Valore	Descrizione
0 (default) (predefinito)	Rende l'ottimizzazione per l'ordinamento delle partizioni dinamiche una decisione basata sui costi quando vengono utilizzati i file ORC. Il numero massimo di scrittori consentiti nelle query INSERT viene calcolato secondo la formula (memoria esecutore/container) * (percentuale di memoria occupata da orc) diviso per la memoria massima (dimensione della striscia occupata da un singolo scrittore).
-1	Disabilita l'ottimizzazione per l'ordinamento completo delle partizioni dinamiche.
1	Consente l'ordinamento globale delle partizioni dinamiche. Ciò mantiene aperto solo un scrittore di record per ogni valore di partizione nel riduttore, riducendo così la pressione della memoria sui riduttori.
2 (or greater integer) (o numero intero maggiore)	Indica a Hive di utilizzare il numero intero specificato come soglia per il numero massimo di scrittori.

Amazon EMR 6.6.0: problemi noti di Hive

- Le query con funzioni finestra sulla stessa colonna di join possono portare a trasformazioni non valide, come riportato in [HIVE-25278](#) e causare risultati errati o errori di query. Come soluzione alternativa, è possibile disabilitare CBO a livello di query per tali query. Contatta il Supporto AWS per ulteriori informazioni.
- Amazon EMR 6.6.0 include la versione 3.1.2 del software Hive. Hive 3.1.2 introduce una caratteristica che divide i file di testo se contengono un'intestazione e un piè di pagina ([HIVE-21924](#)). Apache Tez App Master legge ciascuno dei file per determinare i punti di offset nell'intervallo di dati. La combinazione di questi comportamenti potrebbe influire negativamente sulle prestazioni se le query leggono un numero elevato di file di testo di piccole dimensioni. Come soluzione alternativa, è possibile utilizzare `CombineHiveInputFormat` e regolare la dimensione massima della divisione configurando le seguenti proprietà:

```
SET hive.tez.input.format=org.apache.hadoop.hive ql.io.CombineHiveInputFormat;  
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- Con le versioni da 6.6.0 a 6.9.x di Amazon EMR, le query INSERT con partizione dinamica e una clausola ORDER BY o SORT BY avranno sempre due riduttori. Questo problema è causato dalla modifica dell'OSS [HIVE-20703](#), che assoggetta l'ottimizzazione dinamica delle partizioni di ordinamento a una decisione basata sui costi. Se il carico di lavoro non richiede l'ordinamento delle partizioni dinamiche, ti consigliamo di impostare la proprietà `hive.optimize.sort.dynamic.partition.threshold` su -1 per disabilitare la nuova caratteristica e ottenere il numero di riduttori calcolato correttamente. Questo problema è stato corretto in OSS Hive come parte di [HIVE-22269](#) ed è stato risolto in Amazon EMR 6.10.0.

Hudi

[Apache Hudi](#) è un framework di gestione dei dati open source utilizzato per semplificare l'elaborazione dei dati incrementale e lo sviluppo della pipeline di dati fornendo funzionalità di inserimento, aggiornamento, upsert ed eliminazione a livello di record. Upsert si riferisce alla possibilità di inserire record in un set di dati esistente se non esistono già o di aggiornarli. Gestendo in modo efficiente il modo in cui i dati sono disposti in Amazon S3, Hudi consente di inserire e aggiornare i dati quasi in tempo reale. Hudi mantiene attentamente i metadati delle azioni eseguite sul set di dati per garantire che le azioni siano atomiche e coerenti.

Hudi è integrato con [Apache Spark](#), [Apache Hive](#) e [Presto](#). Nelle versioni Amazon EMR 6.1.0 e successive, Hudi è integrato anche con [Trino \(PrestoSQL\)](#).

Con la versione 5.28.0 e successive di Amazon EMR, EMR installa i componenti Hudi per impostazione predefinita quando sono installati Spark, Hive, Presto o Flink. È possibile utilizzare Spark o l'utilità Hudi DeltaStreamer per creare o aggiornare set di dati Hudi. È possibile utilizzare Hive, Spark, Presto o Flink per eseguire una query a un set di dati Hudi in modo interattivo o creare pipeline di elaborazione dati utilizzando pull incrementali. Il pull incrementale si riferisce alla possibilità di estrarre solo i dati che sono cambiati tra due azioni.

Queste caratteristiche rendono Hudi adatto per i seguenti casi d'uso:

- Utilizzo di dati in streaming da sensori e altri dispositivi IoT (Internet of Things) che richiedono specifici eventi di inserimento e aggiornamento dei dati.
- Rispetto delle norme sulla privacy dei dati nelle applicazioni in cui gli utenti possono scegliere di essere dimenticati o modificare il loro consenso per l'utilizzo dei loro dati.
- Implementazione di un [sistema di acquisizione dati di modifica \(CDC\)](#) che consente di applicare modifiche a un set di dati nel tempo.

La seguente tabella indica la versione di Hudi inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Hudi.

Per la versione dei componenti installati con Hudi in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hudi per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Hudi	Componenti installati con Hudi
emr-6.14.0	Hudi 0.13.1-amzn-2	Not available.

Note

Il rilascio 6.8.0 di Amazon EMR include [Apache Hudi](#) 0.11.1; tuttavia, i cluster Amazon EMR 6.8.0 sono compatibili anche con l'open source `hudi-spark3.3-bundle_2.12` di Hudi 0.12.0.

La seguente tabella indica la versione di Hudi inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Hudi.

Per la versione dei componenti installati con Hudi in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hudi per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Hudi	Componenti installati con Hudi
emr-5.36.1	Hudi 0.10.1-amzn-1	Not available.

Argomenti

- [Funzionamento di Hudi](#)
- [Considerazioni e limitazioni per l'utilizzo di Hudi su Amazon EMR](#)
- [Creazione di un cluster con Hudi installato](#)
- [Lavorare con un set di dati Hudi](#)
- [Utilizzo della CLI Hudi](#)
- [Cronologia dei rilasci di Hudi](#)

Funzionamento di Hudi

Quando si utilizza Hudi con Amazon EMR, è possibile scrivere dati nel set di dati utilizzando l'API Spark Data Source o l'utilità Hudi DeltaStreamer. Hudi organizza un set di dati in una struttura di directory partizionata sotto un *basepath* che è simile a una tradizionale tabella Hive. Le specifiche di come i dati vengono disposti come file in queste directory dipendono dal tipo di set di dati scelto. È possibile scegliere Copia in scrittura (Cow) o Unisci in lettura (mOr).

Indipendentemente dal tipo di set di dati, ogni partizione in un set di dati è identificata in modo univoco dal *partitionpath* relativo al *basepath*. All'interno di ogni partizione, i record vengono distribuiti in più file di dati. Per ulteriori informazioni, consulta [Gestione file](#) nella documentazione di Apache Hudi.

Ogni operazione in Hudi ha un commit corrispondente, identificato da un timestamp crescente in maniera monotona noto come Instant. Hudi dispone tutte le operazioni eseguite sul set di dati in serie su una timeline. Hudi si basa sulla timeline per fornire l'isolamento degli snapshot tra lettori e scrittori e per abilitare il rollback a un punto precedente. Per ulteriori informazioni sulle operazioni che Hudi registra e sul relativo stato, consulta [Timeline](#) nella documentazione di Apache Hudi.

Informazioni sui tipi di archiviazione del set di dati: copia in scrittura rispetto a unisci in lettura

Quando si crea un set di dati Hudi, si specifica che il set di dati è copiato in scrittura o unito in lettura.

- Copia in scrittura (CoW): i dati vengono memorizzati in un formato colonnare (Parquet) e ogni aggiornamento crea una nuova versione dei file durante una scrittura. CoW è il tipo di archiviazione predefinito.
- Unisci in lettura (MoR): i dati vengono memorizzati utilizzando una combinazione di formati colonnare (Parquet) e basati su righe (Avro). Gli aggiornamenti vengono registrati nei file delta basati su righe e vengono compattati in base alle necessità per creare nuove versioni dei file colonnari.

Con i set di dati CoW, ogni volta che c'è un aggiornamento a un record, il file che contiene il record viene riscritto con i valori aggiornati. Quando si lavora con un set di dati MoR, ogniqualvolta è disponibile un aggiornamento Hudi scrive solo la riga per il registro modificato. MoR è più adatto per carichi di lavoro pesanti in scrittura o modifiche con meno letture. CoW è più adatto per carichi di lavoro pesanti di lettura su dati che cambiano meno frequentemente.

Hudi fornisce tre viste logiche per l'accesso ai dati:

- Visualizzazione ottimizzata in lettura: fornisce l'ultimo set di dati con commit delle tabelle CoW e l'ultimo set di dati compattati dalle tabelle MoR.
- Visualizzazione incrementale: fornisce un flusso di modifiche tra due operazioni da un set di dati CoW per alimentare i processi a valle ed estrarre, trasformare e caricare (ETL) flussi di lavoro.
- Visualizzazione in tempo reale: fornisce i dati di commit più recenti da una tabella MoR unendo i file colonnari e basati su righe in linea.

Quando si esegue una query sulla visualizzazione ottimizzata di lettura, la query restituisce tutti i dati compattati ma non include i commit delta più recenti. Eseguire una query di questi dati offre buone prestazioni di lettura, ma omette i dati più aggiornati. Quando si esegue una query sulla visualizzazione in tempo reale, Hudi unisce i dati compattati con i commit delta in lettura. I dati più aggiornati sono disponibili per eseguire query, ma il sovraccarico di calcolo dell'unione rende la query meno performante. La possibilità di eseguire una query dei dati compattati o dati in tempo reale consente di scegliere tra prestazioni e flessibilità quando si esegue una query.

Per ulteriori informazioni sui compromessi tra i tipi di archiviazione, consulta [Tipi di archiviazione e visualizzazioni](#) nella documentazione di Apache Hudi.

Hudi crea due tabelle nel metastore Hive per MoR: una tabella con il nome specificato, che è una vista ottimizzata per la lettura e una tabella con lo stesso nome aggiunto con `_rt`, che è una vista in tempo reale. È possibile eseguire una query a entrambe le tabelle.

Registrazione di un set di dati Hudi con il metastore

Quando si registra una tabella Hudi con il metastore Hive, è possibile interrogare le tabelle Hudi utilizzando Hive, Spark SQL o Presto come si farebbe con qualsiasi altra tabella. Inoltre, è possibile integrare Hudi con AWS Glue configurando Hive e Spark per l'utilizzo di AWS Glue Data Catalog come metastore. Per le tabelle MoR, Hudi registra il set di dati come due tabelle nel Metastore: una tabella con il nome specificato, che è una vista ottimizzata per la lettura e una tabella con lo stesso nome aggiunto con `_rt`, che è una vista in tempo reale.

Viene registrata una tabella Hudi con il metastore Hive quando si utilizza Spark per creare un set di dati Hudi impostando l'opzione `HIVE_SYNC_ENABLED_OPT_KEY` su `"true"` e fornendo altre proprietà richieste. Per ulteriori informazioni, consulta [Lavorare con un set di dati Hudi](#). Inoltre, è possibile utilizzare l'utilità della riga di comando `hive_sync_tool` per registrare un set di dati Hudi come tabella nel metastore, separatamente.

Considerazioni e limitazioni per l'utilizzo di Hudi su Amazon EMR

- Il campo della chiave di registro non può essere null o vuoto: il campo specificato come chiave di registro non può avere valori null o vuoti.
- Schema aggiornato per impostazione predefinita su upsert e insert: Hudi fornisce un'interfaccia, `HoodieRecordPayload`, che determina il modo in cui il `DataFrame` di input e il set di dati Hudi esistente vengono uniti per produrre un nuovo set di dati aggiornato. Hudi fornisce un'implementazione predefinita di questa classe, `OverwriteWithLatestAvroPayload`, che sovrascrive i registri esistenti e aggiorna lo schema come specificato nel `DataFrame` di input. Per personalizzare questa logica per l'implementazione di aggiornamenti di unione e parziali, è possibile fornire un'implementazione dell'interfaccia `HoodieRecordPayload` utilizzando il parametro `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY`.
- L'eliminazione richiede lo schema: durante l'eliminazione, è necessario specificare la chiave di registro, la chiave di partizione e i campi chiave pre-combinati. Altre colonne possono essere rese null o vuote, ma è necessario lo schema completo.
- Limitazioni delle tabelle MoR: le tabelle MoR non supportano il savepointing. È possibile interrogare le tabelle MoR utilizzando la vista ottimizzata per la lettura o la vista in tempo reale (`tableName_rt`) da Spark SQL, Presto o Hive. L'utilizzo della visualizzazione ottimizzata per la lettura espone solo i dati del file di base e non espone una vista unita dei dati di base e di log.
- Hive
 - Per la registrazione delle tabelle nel metastore Hive, Hudi si aspetta che il server Hive Thrift sia in esecuzione sulla porta predefinita 10000. Se si esegue l'override di questa porta con una porta personalizzata, passa l'opzione `HIVE_URL_OPT_KEY` come illustrato nell'esempio seguente.

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number")
```

- Il tipo di dati `timestamp` in Spark è registrato come tipo di dati `long` in Hive e non come tipo `timestamp` di Hive.
- Presto
 - Presto non supporta la lettura di tabelle MoR in tempo reale nelle versioni Hudi precedenti alla 0.6.0.
 - Presto supporta solo le query degli snapshot.

- Affinché Presto interpreti correttamente le colonne Hudi del set di dati, imposta il valore `hive.parquet_use_column_names` su `true`.
- Per impostare il valore di una sessione, nella shell Presto, eseguire il comando seguente:

```
set session hive.parquet_use_column_names=true
```

- Per impostare il valore a livello di cluster, utilizzare la classificazione di configurazione `presto-connector-hive` per impostare `hive.parquet.use_column_names` su `true`, come illustrato nell'esempio seguente. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.parquet.use-column-names": "true"
    }
  }
]
```

- Indice HBase
- La versione di HBase utilizzata per costruire Hudi potrebbe essere diversa da quella indicata nella Guida ai rilasci di EMR. Per inserire le dipendenze corrette per la sessione Spark, esegui il comando seguente.

```
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"
```

Creazione di un cluster con Hudi installato

Con la versione 5.28.0 e successive di Amazon EMR, Amazon EMR installa i componenti Hudi per impostazione predefinita quando sono installati Spark, Hive o Presto. Per utilizzare Hudi su Amazon EMR, crea un cluster con una o più delle seguenti applicazioni installate:

- Hadoop
- Hive

- Spark
- Presto
- Flink

Puoi creare un cluster utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR.

Creazione di un cluster con Hudi utilizzando la AWS Management Console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Configurazione software, scegli emr-5.28.0 o versioni successive per Release (Rilascio) e seleziona Hadoop, Hive, Spark, Presto e Tez insieme ad altre applicazioni richieste dal cluster.
4. Configurare altre opzioni come richiesto per l'applicazione, quindi scegliere Avanti.
5. Configurare le opzioni per Hardware e Impostazioni generali del cluster come desiderato.
6. Per Opzioni di protezione, si consiglia di selezionare una coppia di chiavi EC2 che è possibile utilizzare per connettersi alla riga di comando del nodo master utilizzando SSH. In questo modo è possibile eseguire i comandi della shell Spark, i comandi della CLI Hive e i comandi della CLI Hudi descritti in questa guida.
7. Configurare altre opzioni di sicurezza come desiderato, quindi scegliere Crea cluster.

Lavorare con un set di dati Hudi

Hudi supporta l'inserimento, l'aggiornamento e l'eliminazione di dati nei set di dati Hudi tramite Spark. Per ulteriori informazioni, consulta [Scrittura delle tabelle Hudi](#) nella documentazione di Apache Hudi.

Negli esempi seguenti viene illustrato come avviare la shell interattiva Spark, utilizzare Spark submit o utilizzare Amazon EMR Notebooks per lavorare con Hudi su Amazon EMR. È inoltre possibile utilizzare l'utilità Hudi DeltaStreamer o altri strumenti per scrivere su un set di dati. In questa sezione, gli esempi illustrano l'utilizzo di set di dati con la shell Spark connessi al nodo master utilizzando SSH come utente hadoop predefinito.

Avvio della shell (interprete di comandi) Spark utilizzando Amazon EMR versione 6.7 e successive

Durante l'esecuzione di `spark-shell`, `spark-submit` o `spark-sql` utilizzando Amazon EMR versione 6.7.0 o successive, esegui i seguenti comandi.

Note

Amazon EMR 6.7.0 impiega [Apache Hudi 0.11.0-amzn-0](#), che contiene miglioramenti significativi rispetto alle versioni precedenti di Hudi. Per ulteriori informazioni, consulta la sezione [Apache Hudi 0.11.0 Migration Guide](#) (Guida alla migrazione di Apache Hudi 0.11.0). Gli esempi in questa scheda riflettono tali modifiche.

Per aprire la shell (interprete di comandi) Spark nel nodo primario

1. Effettua la connessione al nodo primario tramite SSH. Per ulteriori informazioni, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida alla gestione di Amazon EMR.
2. Immettere il seguente comando per avviare la shell Spark. Per utilizzare la shell PySpark, sostituisci `spark-shell` con `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"
```

Avvio della shell (interprete di comandi) Spark utilizzando Amazon EMR 6.6 e rilasci precedenti

Durante l'esecuzione di `spark-shell`, `spark-submit` o `spark-sql` utilizzando Amazon EMR 6.6.x o rilasci precedenti, esegui i seguenti comandi.

Note

- Amazon EMR versioni 6.2 e 5.31 e successive (Hudi versione 0.6.x e successive) può omettere `spark-avro.jar` dalla configurazione.
- Amazon EMR rilasci 6.5 e 5.35 e successivi (Hudi rilascio 0.9.x e successivi) può omettere `spark.sql.hive.convertMetastoreParquet=false` dalla configurazione.
- Amazon EMR versioni 6.6 e 5.36 e successive (Hudi versione 0.10.x e successive) deve includere la configurazione `HoodieSparkSessionExtension` descritta in [Version: 0.10.0 Spark Guide](#) (Guida a Spark versione 0.10.0):

```
--conf  
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"  
\
```

Per aprire la shell (interprete di comandi) Spark nel nodo primario

1. Effettua la connessione al nodo primario tramite SSH. Per ulteriori informazioni, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida alla gestione di Amazon EMR.
2. Immettere il seguente comando per avviare la shell Spark. Per utilizzare la shell PySpark, sostituisci `spark-shell` con `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
```

Utilizzo di Hudi con i notebook Amazon EMR utilizzando Amazon EMR versione 6.7 e successive

Per utilizzare Hudi con Amazon EMR Notebooks, è necessario innanzitutto copiare i file jar Hudi dal file system locale in HDFS sul nodo master del cluster di notebook. È quindi possibile utilizzare l'editor del notebook per configurare EMR Notebooks per l'utilizzo di Hudi.

Utilizzo di Hudi con Amazon EMR Notebooks

1. Crea e avvia un cluster per Amazon EMR Notebooks. Per ulteriori informazioni, consulta [Creazione di cluster per i notebook Amazon EMR](#) nella Guida alla gestione di Amazon EMR.
2. Connettersi al nodo master del cluster utilizzando SSH e quindi copiare i file jar dal filesystem locale in HDFS, come illustrato negli esempi seguenti. Nell'esempio, creiamo una directory in HDFS per chiarezza nella gestione dei file. È possibile scegliere la propria destinazione in HDFS, se lo si desidera.

```
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. Aprire l'editor del notebook, immettere il codice dall'esempio seguente ed eseguirlo.

```
%%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"
  }}
}
```

Utilizzo di Hudi con i notebook Amazon EMR utilizzando Amazon EMR 6.6 e rilasci precedenti

Per utilizzare Hudi con Amazon EMR Notebooks, è necessario innanzitutto copiare i file jar Hudi dal file system locale in HDFS sul nodo master del cluster di notebook. È quindi possibile utilizzare l'editor del notebook per configurare EMR Notebooks per l'utilizzo di Hudi.

Utilizzo di Hudi con Amazon EMR Notebooks

1. Crea e avvia un cluster per Amazon EMR Notebooks. Per ulteriori informazioni, consulta [Creazione di cluster per i notebook Amazon EMR](#) nella Guida alla gestione di Amazon EMR.

2. Connettersi al nodo master del cluster utilizzando SSH e quindi copiare i file jar dal filesystem locale in HDFS, come illustrato negli esempi seguenti. Nell'esempio, creiamo una directory in HDFS per chiarezza nella gestione dei file. È possibile scegliere la propria destinazione in HDFS, se lo si desidera.

```
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. Aprire l'editor del notebook, immettere il codice dall'esempio seguente ed eseguirlo.

```
{ "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"  
}}
```

Inizializzazione di una sessione Spark per Hudi

Quando si utilizza Scala, è necessario importare le seguenti classi nella sessione Spark. Questo deve essere fatto una volta per sessione Spark.

```
import org.apache.spark.sql.SaveMode  
import org.apache.spark.sql.functions._  
import org.apache.hudi.DataSourceWriteOptions  
import org.apache.hudi.DataSourceReadOptions  
import org.apache.hudi.config.HoodieWriteConfig  
import org.apache.hudi.hive.MultiPartKeysValueExtractor  
import org.apache.hudi.hive.HiveSyncConfig  
import org.apache.hudi.sync.common.HoodieSyncConfig
```

Scrittura in un set di dati Hudi

Negli esempi seguenti viene illustrato come creare un DataFrame e scriverlo come set di dati Hudi.

Note

Per incollare gli esempi di codice nella shell Spark, digitare **:paste** al prompt, incollare l'esempio e premere **CTRL + D**.

Ogni volta che si scrive un DataFrame in un set di dati Hudi, è necessario specificare `DataSourceWriteOptions`. Molte di queste opzioni sono probabilmente identiche tra le operazioni di scrittura. L'esempio seguente specifica le opzioni comuni utilizzando la variabile `hudiOptions`, che gli esempi successivi utilizzano.

Scrittura utilizzando Scala con Amazon EMR versione 6.7 e successive

Note

Amazon EMR 6.7.0 impiega [Apache Hudi 0.11.0-amzn-0](#), che contiene miglioramenti significativi rispetto alle versioni precedenti di Hudi. Per ulteriori informazioni, consulta la sezione [Apache Hudi 0.11.0 Migration Guide](#) (Guida alla migrazione di Apache Hudi 0.11.0). Gli esempi in questa scheda riflettono tali modifiche.

```
// 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/"))

```

Scrittura utilizzando Scala con Amazon EMR 6.6 e rilasci precedenti

```

// 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/"))
```

Scrittura usando PySpark

```
# Create a DataFrame
inputDF = spark.createDataFrame(
  [
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
    ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
    ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
  ],
  ["id", "creation_date", "last_update_time"]
)

# Specify common DataSourceWriteOptions in the single hudiOptions variable
hudiOptions = {
  'hoodie.table.name': 'tableName',
  'hoodie.datasource.write.recordkey.field': 'id',
  'hoodie.datasource.write.partitionpath.field': 'creation_date',
  'hoodie.datasource.write.precombine.field': 'last_update_time',
  'hoodie.datasource.hive_sync.enable': 'true',
  'hoodie.datasource.hive_sync.table': 'tableName',
  'hoodie.datasource.hive_sync.partition_fields': 'creation_date',
  'hoodie.datasource.hive_sync.partition_extractor_class':
  'org.apache.hudi.hive.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

Potresti visualizzare "hoodie" invece di Hudi negli esempi di codice e nelle notifiche. La base di codice Hudi utilizza ampiamente la vecchia ortografia "hoodie".

Riferimento DataSourceWriteOptions per Hudi

Opzione	Descrizione
TABLE_NAME	Il nome della tabella in cui registrare il set di dati.
TABLE_TYPE_OPT_KEY	Facoltativo. Specificare se il set di dati viene creato come "COPY_ON_WRITE" o "MERGE_ON_READ". Il valore predefinito è "COPY_ON_WRITE".
RECORDKEY_FIELD_OPT_KEY	Il campo chiave record il cui valore verrà utilizzato come componente <code>recordKey</code> di <code>HoodieKey</code> . Il valore effettivo sarà ottenuto richiamando <code>toString()</code> sul valore del campo. I campi nidificati possono essere specificati utilizzando la notazione con punti, ad esempio <code>a.b.c</code> .
PARTITIONPATH_FIELD_OPT_KEY	Il campo del percorso di partizione il cui valore verrà utilizzato come componente <code>partitionPath</code> di <code>HoodieKey</code> . Il valore effettivo sarà ottenuto richiamando <code>toString()</code> sul valore del campo.
PRECOMBINE_FIELD_OPT_KEY	Il campo utilizzato nella pre-combinazione prima della scrittura effettiva. Quando due registri hanno lo stesso valore chiave, Hudi seleziona quello con il valore più grande per il

Opzione	Descrizione
	campo di precombinazione come determinato da <code>Object.compareTo(..)</code> .

Le seguenti opzioni sono necessarie solo per registrare la tabella del set di dati Hudi nel metastore. Se non si registra il set di dati Hudi come tabella nel metastore Hive, queste opzioni non sono necessarie.

Riferimento DataSourceWriteOptions per Hive

Opzione	Descrizione
HIVE_DATABASE_OPT_KEY	Il database Hive da sincronizzare. Il valore predefinito è "default" .
HIVE_PARTITION_EXTRACTOR_CLASSES_OPT_KEY	La classe utilizzata per estrarre i valori dei campi delle partizioni nelle colonne delle partizioni Hive.
HIVE_PARTITION_FIELDS_OPT_KEY	Campo nel set di dati da utilizzare per determinare le colonne di partizione Hive.
HIVE_SYNC_ENABLED_OPT_KEY	Quando impostato su "true", registra il set di dati con il metastore Apache Hive. Il valore predefinito è "false".
HIVE_TABLE_OPT_KEY	Campo obbligatorio. Il nome della tabella in Hive da sincronizzare. Ad esempio, "my_hudi_table_cow" .
HIVE_USER_OPT_KEY	Facoltativo. Il nome utente Hive da utilizzare e durante la sincronizzazione. Ad esempio, "hadoop".
HIVE_PASS_OPT_KEY	Facoltativo. La password Hive per l'utente specificato da HIVE_USER_OPT_KEY .
HIVE_URL_OPT_KEY	L'URL del metastore Hive.

Upsert dei dati

Nell'esempio seguente viene illustrato come eseguire l'upsert dei dati scrivendo un DataFrame. A differenza dell'esempio di inserimento precedente, il valore `OPERATION_OPT_KEY` è impostato su `UPSERT_OPERATION_OPT_VAL`. Inoltre, `.mode(SaveMode.Append)` è specificato per indicare che il record deve essere aggiunto.

Upsert utilizzando Scala con Amazon EMR rilascio 6.7 e successivi

Note

Amazon EMR 6.7.0 impiega [Apache Hudi 0.11.0-amzn-0](#), che contiene miglioramenti significativi rispetto alle versioni precedenti di Hudi. Per ulteriori informazioni, consulta la sezione [Apache Hudi 0.11.0 Migration Guide](#) (Guida alla migrazione di Apache Hudi 0.11.0). Gli esempi in questa scheda riflettono tali modifiche.

```
// 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 utilizzando Scala con Amazon EMR 6.6 e rilasci precedenti

```
// 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 utilizzando 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/')
```

Eliminazione di un registro

Per eliminare un registro, è possibile eseguire l'upsert di un payload vuoto. In questo caso, l'opzione `PAYLOAD_CLASS_OPT_KEY` specifica la classe `EmptyHoodieRecordPayload`. Nell'esempio viene utilizzato lo stesso DataFrame, `updateDF`, utilizzato nell'esempio upsert per specificare lo stesso registro.

Eliminazione utilizzando Scala con Amazon EMR rilascio 6.7 e successivi

Note

Amazon EMR 6.7.0 impiega [Apache Hudi 0.11.0-amzn-0](#), che contiene miglioramenti significativi rispetto alle versioni precedenti di Hudi. Per ulteriori informazioni, consulta la sezione [Apache Hudi 0.11.0 Migration Guide](#) (Guida alla migrazione di Apache Hudi 0.11.0). Gli esempi in questa scheda riflettono tali modifiche.

```
(updateDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
```

```
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

Eliminazione utilizzando Scala con Amazon EMR 6.6 e rilasci precedenti

```
(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/"))
```

Eliminazione utilizzando 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/')
```

È inoltre possibile eliminare i dati impostando `OPERATION_OPT_KEY` su `DELETE_OPERATION_OPT_VAL` per rimuovere tutti i registri nel set di dati inviato. Per istruzioni sull'esecuzione delle eliminazioni soft e per ulteriori informazioni sull'eliminazione dei dati archiviati nelle tabelle Hudi, consulta [Deletes \(Eliminazioni\)](#) nella documentazione di Apache Hudi.

Lettura da un set di dati Hudi

Per recuperare i dati al momento attuale, Hudi esegue query degli snapshot per impostazione predefinita. Di seguito è riportato un esempio di eseguire una query del set di dati scritto su S3 in [Scrittura in un set di dati Hudi](#). Sostituisci `s3://DOC-EXAMPLE-BUCKET/myhuidataset` con il percorso della tabella e aggiungi asterischi con caratteri jolly per ogni livello di partizione, più un asterisco aggiuntivo. Dal momento che questo esempio mostra un livello di partizione, sono stati aggiunti due simboli jolly.

Lettura utilizzando Scala con Amazon EMR rilascio 6.7 e successivi

Note

Amazon EMR 6.7.0 impiega [Apache Hudi 0.11.0-amzn-0](#), che contiene miglioramenti significativi rispetto alle versioni precedenti di Hudi. Per ulteriori informazioni, consulta la sezione [Apache Hudi 0.11.0 Migration Guide](#) (Guida alla migrazione di Apache Hudi 0.11.0). Gli esempi in questa scheda riflettono tali modifiche.

```
(val snapshotQueryDF = spark.read
  .format("hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset")
  .show())
```

Lettura utilizzando Scala con Amazon EMR 6.6 e rilasci precedenti

```
(val snapshotQueryDF = spark.read
  .format("org.apache.hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

Lettura utilizzando PySpark

```
snapshotQueryDF = spark.read \
  .format('org.apache.hudi') \
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')

snapshotQueryDF.show()
```

Query incrementali

È inoltre possibile eseguire query incrementali con Hudi per ottenere un flusso di registri che hanno subito modifiche a partire da un dato timestamp di commit. Per fare questo, imposta il campo `QUERY_TYPE_OPT_KEY` su `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Quindi, aggiungi un valore per `BEGIN_INSTANTTIME_OPT_KEY` per ottenere tutti i registri scritti dall'ora specificata. Le query incrementali sono in genere dieci volte più efficienti rispetto alle controparti batch poiché elaborano solo i registri modificati.

Quando si eseguono query incrementali, utilizza il percorso della tabella radice (di base) senza gli asterischi jolly utilizzati per le query degli snapshot.

Note

Presto non supporta le query incrementali.

Query incrementali utilizzando 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()
```

Query incrementali utilizzando 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()
```

Per ulteriori informazioni sulla lettura dai set di dati Hudi, consulta [Esecuzione di query sulle tabelle Hudi](#) nella documentazione di Apache Hudi.

Utilizzo della CLI Hudi

È possibile utilizzare la CLI Hudi per gestire i set di dati Hudi e visualizzare informazioni su commit, file system, statistiche e altro ancora. È inoltre possibile utilizzare la CLI per eseguire manualmente

le compattazioni, pianificare le compattazioni o annullare le compattazioni pianificate. Per ulteriori informazioni, consulta [Interazione tramite CLI](#) nella documentazione di Apache Hudi.

Avvio della CLI Hudi e connessione a un set di dati

1. Connessione al nodo master tramite SSH Per ulteriori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
2. Nella riga di comando, digita `/usr/lib/hudi/cli/bin/hudi-cli.sh`.

Il prompt dei comandi viene modificato in `hudi->`.

3. Digitare il codice seguente per connettersi a un set di dati. Sostituire `s3://DOC-EXAMPLE-BUCKET/myhudidataset` con il percorso del set di dati con cui si desidera lavorare. Il valore che usiamo è lo stesso del valore stabilito negli esempi precedenti.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhudidataset
```

Il prompt dei comandi cambia per includere il set di dati a cui si è connessi, come illustrato nell'esempio seguente.

```
hudi:myhudidataset->
```

Cronologia dei rilasci di Hudi

La seguente tabella indica la versione di Hudi inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Hudi

Etichetta di rilascio di Amazon EMR	Versione di Hudi	Componenti installati con Hudi
emr-6.14.0	0.13.1-amzn-2	Not available.
emr-6.13.0	0.13.1-amzn-1	Not available.
emr-6.12.0	0.13.1-amzn-0	Not available.

Etichetta di rilascio di Amazon EMR	Versione di Hudi	Componenti installati con Hudi
emr-6.11.1	0.13.0-amzn-0	Not available.
emr-6.11.0	0.13.0-amzn-0	Not available.
emr-6.10.1	0.12.2-amzn-0	Not available.
emr-6.10.0	0.12.2-amzn-0	Not available.
emr-6.9.1	0.12.1-amzn-0	Not available.
emr-6.9.0	0.12.1-amzn-0	Not available.
emr-6.8.1	0.11.1-amzn-0	Not available.
emr-6.8.0	0.11.1-amzn-0	Not available.
emr-6.7.0	0.11.0-amzn-0	Not available.
emr-5.36.1	0.10.1-amzn-1	Not available.
emr-5.36.0	0.10.1-amzn-1	Not available.
emr-6.6.0	0.10.1-amzn-0	Not available.
emr-5.35.0	0.9.0-amzn-2	Not available.
emr-6.5.0	0.9.0-amzn-1	Not available.
emr-6.4.0	0.8.0-amzn-0	Not available.
emr-6.3.1	0.7.0-amzn-0	Not available.
emr-6.3.0	0.7.0-amzn-0	Not available.
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.

Etichetta di rilascio di Amazon EMR	Versione di Hudi	Componenti installati con Hudi
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.
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) è un'interfaccia utente grafica, basata sul Web, open source per l'utilizzo con Amazon EMR e Apache Hadoop. Hue consente di raggruppare diversi progetti di ecosistema Hadoop in un'interfaccia configurabile. Amazon EMR ha inoltre aggiunto personalizzazioni specifiche per Hue in Amazon EMR. Hue funge da front-end per applicazioni in esecuzione sul cluster, consentendo di interagire con applicazioni utilizzando un'interfaccia che potrebbe risultare più familiare o descrittiva. Le applicazioni in Hue, ad esempio gli editor Hive e Pig, sostituiscono l'esigenza di accedere al cluster per eseguire script in modo interattivo utilizzando la rispettiva shell di ogni applicazione. Dopo l'avvio di un cluster, puoi interagire completamente con applicazioni utilizzando Hue o un'interfaccia simile. Per ulteriori informazioni su Hue, consulta <http://gethue.com>.

Hue viene installato per impostazione predefinita all'avvio del cluster utilizzando la console di Amazon EMR. Puoi scegliere di non installare Hue utilizzando Advanced options (Opzioni avanzate) nella console di Amazon EMR all'avvio di un cluster, oppure specificando in modo esplicito l'opzione `--applications` e omettendo Hue, quando utilizzi `create-cluster` da AWS CLI.

La seguente tabella indica la versione di Hue inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Hue.

Per la versione dei componenti installati con Hue in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hue per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-6.14.0	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,

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
		mariadb-server, oozie-client, oozie-server

La seguente tabella indica la versione di Hue inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Hue.

Per la versione dei componenti installati con Hue in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Hue per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.36.1	Hue 4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Argomenti

- [Caratteristiche supportate e non supportate di Hue su Amazon EMR](#)
- [Connessione all'interfaccia utente Web di Hue](#)
- [Utilizzo di Hue con un database remoto in Amazon RDS](#)
- [Configurazioni avanzate per Hue](#)
- [Cronologia dei rilasci di Hue](#)

Caratteristiche supportate e non supportate di Hue su Amazon EMR

- Amazon S3 e browser Hadoop File System (HDFS)
 - Con le autorizzazioni appropriate, puoi selezionare e spostare dati tra lo storage HDFS temporaneo e i bucket S3 appartenenti al tuo account.
 - Per impostazione predefinita, i superuser su Hue possono accedere a tutti i file a cui hanno accesso i ruoli IAM di Amazon EMR. Gli utenti appena creati non dispongono automaticamente delle autorizzazioni per accedere al browser di file Amazon S3 e devono disporre delle autorizzazioni `filebrowser.s3_access` attivate per il proprio gruppo.
- Hive consente di eseguire query interattive sui dati. Questo è anche un modo utile per realizzare prototipi di query programmatiche o in batch.
- Pig - Consente di eseguire script sui dati o inviare comandi interattivi.
- Oozie - Consente di creare e monitorare flussi di lavoro Oozie.
- Metastore Manager - Consente di visualizzare e modificare i contenuti del metastore Hive (importare/creare, rilasciare e così via).
- Job browser - Consente di visualizzare lo stato dei processi Hadoop inviati.
- Gestione degli utenti: consente di gestire account utente Hue e integrare utenti LDAP con Hue.
- Esempi AWS: esistono diversi esempi "pronti per l'esecuzione" che elaborano i dati campione di diversi servizi AWS utilizzando le applicazioni disponibili in Hue. Quando si esegue l'accesso a Hue, viene eseguito il reindirizzamento all'applicazione Hue Home in cui i campioni sono preinstallati.
- Livy Server è supportato solo in Amazon EMR versione 5.9.0 e successive.
- Per utilizzare Hue Notebook for Spark, devi installare Hue con Livy e Spark.
- Il pannello di controllo Hue non è supportato.
- PostgreSQL non è supportato.

Connessione all'interfaccia utente Web di Hue

Connettersi all'interfaccia utente Web di Hue equivale a connettersi a qualsiasi interfaccia HTTP ospitata sul nodo master di un cluster. Nella procedura seguente viene descritto come accedere all'interfaccia utente di Hue. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce Web ospitate nei cluster Amazon EMR](#) nella Guida per la gestione di Amazon EMR.

Visualizzazione dell'interfaccia utente Web di Hue

1. Segui queste istruzioni per [Configurare un tunnel SSH sul nodo master utilizzando l'inoltro porta dinamico](#) nella Guida per la gestione di Amazon EMR.
2. Digitare il seguente indirizzo nel browser per aprire l'interfaccia Web Hue: `http://master public DNS:8888` in cui *master public dns* è il nome DNS pubblico del nodo master del cluster, ad esempio `ec2-11-22-333-44.compute-1.amazonaws.com`.
3. La prima volta che accedi come amministratore, inserisci nella schermata di accesso di Hue un nome utente e una password per creare l'account super Hue, quindi seleziona Crea account. In caso contrario, digitare il nome utente e la password e selezionare Create account (Crea account) o immettere le credenziali fornite dall'amministratore.

Utilizzo di Hue con un database remoto in Amazon RDS

Per impostazione predefinita, le informazioni utente Hue e le cronologie delle query vengono archiviate in un database MySQL locale sul nodo master. In alternativa, puoi creare uno o più cluster abilitati Hue utilizzando una configurazione archiviata in Amazon S3 e un database MySQL in Amazon Relational Database Service (Amazon RDS). In questo modo puoi preservare le informazioni utente e la cronologia delle query create da Hue senza mantenere il cluster Amazon EMR in esecuzione. Ti consigliamo di usare la crittografia lato server Amazon S3 per archiviare il file di configurazione.

Crea innanzitutto il database remoto per Hue.

Creazione del database MySQL esterno

1. Apri la console di Amazon RDS all'indirizzo <https://console.aws.amazon.com/rds/>.
2. Fare clic su Launch a DB Instance (Avvia un'istanza database).
3. Scegliere MySQL e fare clic su Select (Seleziona).
4. Lasciare la selezione predefinita di Multi-AZ Deployment and Provisioned IOPS Storage (Implementazione Multi-AZ e storage IOPS assegnato) e fare clic su Next (Successivo).
5. Lasciare le impostazioni predefinite per Instance Specifications (Specifiche delle istanze), specificare Settings (Impostazioni) e fare clic su Next (Successivo).
6. Nella pagina Configure Advanced Settings (Configura impostazioni avanzate), scegliere un gruppo di sicurezza e un nome del database adeguati. Il gruppo di sicurezza utilizzato deve consentire almeno l'accesso TCP in entrata per porta 3306 dal nodo master del cluster. Se a

questo punto il cluster non è ancora stato creato, puoi consentire a tutti gli host di connettersi alla porta 3306 e modificare il gruppo di sicurezza dopo che il cluster è stato avviato. Fare clic su Launch DB Instance (Avvia istanza database).

7. Dal pannello di controllo RDS, scegliere Instances (Istanze) e selezionare l'istanza appena creata. Quando il database è disponibile, prendere nota di `namedb`, nome utente, password e nome host dell'istanza RDS. Utilizzare questa informazione quando si crea e configura il cluster.

Per specificare un database MySQL esterno per Hue quando si avvia un cluster utilizzando AWS CLI

Per specificare un database MySQL esterno per Hue quando si avvia un cluster utilizzando AWS CLI, utilizzare le informazioni annotate durante la creazione dell'istanza RDS per configurare `hue.ini` con un oggetto di configurazione

Note

Puoi creare più cluster che utilizzano lo stesso database esterno, ma ogni cluster condividerà la cronologia delle query e le informazioni utente.

- Utilizzando AWS CLI, creare un cluster con Hue installato, utilizzando il database esterno creato e facendo riferimento a un file di configurazione con una classificazione della configurazione per Hue che specifica le proprietà del database. L'esempio seguente consente di creare un cluster con Hue installato, facendo riferimento a un file di configurazione in Amazon S3, `myConfig.json`, che specifica la configurazione del database.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --release-label emr-5.36.1 --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
```


Contenuti di esempio del file `myConfig.json` sono mostrati di seguito. Sostituisci *dbname*, *nome utente*, *password* e *nome host dell'istanza RDS* con i valori annotati in precedenza nel pannello di controllo RDS.

```
[{
  "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": []
        }
      ]
    }
  ]
}]
```

Risoluzione dei problemi

In caso di failover di Amazon RDS

Durante l'esecuzione di una query si possono verificare ritardi perché l'istanza del database Hue non risponde o è in corso di esecuzione un failover. Di seguito sono elencati alcuni fatti e linee guida per questo problema:

- Se esegui l'accesso alla console Amazon RDS, puoi cercare gli eventi di failover. Ad esempio, per vedere se un failover è in corso o si è verificato, cerca eventi quali "Multi-AZ instance failover started" e "Multi-AZ instance failover completed".

- Un'istanza RDS impiega circa 30 secondi per completare un failover.
- Se le risposte alle query in Hue richiedono più tempo del previsto, prova a eseguire nuovamente la query.

Configurazioni avanzate per Hue

Questa sezione contiene gli argomenti seguenti.

Argomenti

- [Configurazione di Hue per utenti LDAP](#)

Configurazione di Hue per utenti LDAP

L'integrazione con LDAP consente agli utenti di accedere a Hue utilizzando le credenziali esistenti archiviate in una directory LDAP. Quando si integra Hue con LDAP, non è necessario gestire in modo indipendente le informazioni utente in Hue. Le informazioni riportate di seguito dimostrano l'integrazione Hue con Microsoft Active Directory, ma le opzioni di configurazione sono analoghe a quelle di qualsiasi directory LDAP.

L'autenticazione LDAP esegue innanzitutto l'associazione al server e stabilisce la connessione. Quindi, la connessione stabilita viene utilizzata per qualsiasi query successiva per cercare le informazioni utente LDAP. A meno che il server Active Directory non consenta connessioni anonime, una connessione deve essere stabilita utilizzando nome e password distinti dell'associazione. Il nome distinto dell'associazione (o DN) è definito dall'impostazione di configurazione `bind_dn`. La password dell'associazione è definita dall'impostazione di configurazione `bind_password`. In Hue sono disponibili due modi per associare richieste LDAP: associazione di ricerca e associazione diretta. Il metodo preferito per utilizzare Hue con Amazon EMR è l'associazione di ricerca.

Quando si utilizza l'associazione di ricerca con Active Directory, Hue utilizza l'attributo nome utente (definito da `user_name_attr config`) per trovare l'attributo che deve essere recuperato dal nome distinto di base (o DN). L'associazione di ricerca è utile quando il DN completo per l'utente Hue non è noto.

Ad esempio, puoi impostare `user_name_attr config` per utilizzare il nome comune (o CN). In questo caso, il server Active Directory utilizza il nome utente Hue fornito durante l'accesso per cercare nella struttura di directory un nome comune corrispondente, partendo dal nome distinto di base. Se il nome comune per l'utente Hue viene trovato, il nome distinto dell'utente viene restituito

dal server. Hue costruisce quindi un nome distinto utilizzato per autenticare l'utente eseguendo un'operazione di associazione.

Note

L'associazione di ricerca consente di cercare i nomi utente in tutte le sottostrutture di directory partendo dal nome distinto di base. Il nome distinto di base specificato nella configurazione LDAP Hue deve essere il parente più prossimo del nome utente. In caso contrario, le prestazioni dell'autenticazione LDAP potrebbero peggiorare.

Quando si utilizza l'associazione diretta con Active Directory, utilizzare `nt_domain` o `ldap_username_pattern` esatti per eseguire l'autenticazione. Quando si utilizza l'associazione diretta, se è definito l'attributo di dominio `nt` (definito dall'impostazione di configurazione `nt_domain`), viene creato un modello di nome distinto dell'utente utilizzando il modulo: `<login username>@nt_domain`. Questo modello viene utilizzato per eseguire la ricerca in tutte le sottostrutture di directory partendo dal nome distinto di base. Se il dominio `nt` non è configurato, Hue esegue la ricerca di un modello di nome distinto esatto per l'utente (definito dall'impostazione di configurazione `ldap_username_pattern`). In questo caso, il server cerca un valore `ldap_username_pattern` corrispondente in tutte le sottostrutture di directory partendo dal nome distinto di base.

Avvio di un cluster con proprietà LDAP per Hue utilizzando la AWS CLI

- Per specificare le proprietà LDAP per `hue-ini`, creare un cluster con Hue installato e fare riferimento a un file json con proprietà di configurazione per LDAP. Di seguito è mostrato un comando esempio che fa riferimento a un file di configurazione `myConfig.json` archiviato in Amazon S3.

```
aws emr create-cluster --release-label emr-5.36.1 --applications Name=Hue
Name=Spark Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json.
```

Contenuti di esempio di `myConfig.json` sono mostrati di seguito.

```
[
  {
    "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

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Visualizzazione delle impostazioni LDAP in Hue

1. Accertarsi di disporre di una connessione VPN o di un tunnel SSH attivo al nodo master del cluster Amazon EMR. Quindi, nel browser, digitare *master-public-dns*: 8888 per aprire l'interfaccia Web Hue.
2. Accedere usando le credenziali di amministratore Hue. Se si apre la finestra Did you know? (Lo sapevi?), fare clic su Got it, prof! (Ricevuto, prof!) per chiuderla.
3. Fare clic sull'icona Hue nella barra degli strumenti.
4. Nella pagina About Hue (Informazioni su Hue), fare clic su Configuration (Configurazione).
5. Nella sezione Configuration Sections and Variables (Sezioni e variabili di configurazione), fare clic su Desktop.
6. Scorrere fino alla sezione Idap per visualizzare le impostazioni.

Cronologia dei rilasci di Hue

La tabella seguente indica la versione di Hue inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x o Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Hue

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-6.14.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode,

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
		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-6.13.0	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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.33.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.33.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.31.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.31.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.28.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.28.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.26.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.25.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.24.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.24.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.23.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.23.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.22.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.18.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.18.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.17.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.17.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.16.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.16.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.15.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.14.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.13.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.13.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.12.3	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.12.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.11.4	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.3	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.11.2	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.11.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.10.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.10.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.9.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.8.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.8.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-5.0.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.9.6	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-4.9.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-4.9.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-4.8.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.0	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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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-4.7.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
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-4.6.0	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

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-4.4.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.3.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.2.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Etichetta di rilascio di Amazon EMR	Versione di Hue	Componenti installati con Hue
emr-4.1.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Iceberg

[Apache Iceberg](#) è un formato di tabella aperta per serie di dati di grandi dimensioni in Amazon Simple Storage Service (Amazon S3). Fornisce prestazioni di query rapide su tabelle di grandi dimensioni, commit atomici, scritture simultanee ed evoluzione delle tabelle compatibili con SQL. A partire da Amazon EMR 6.5.0, puoi utilizzare Apache Spark 3 su cluster Amazon EMR con il formato tabella Iceberg.

La seguente tabella indica la versione di Iceberg inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Iceberg.

Per la versione dei componenti installati con Iceberg in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Iceberg per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Iceberg	Componenti installati con Iceberg
emr-6.14.0	Iceberg 1.3.1-amzn-0	Not available.

Argomenti

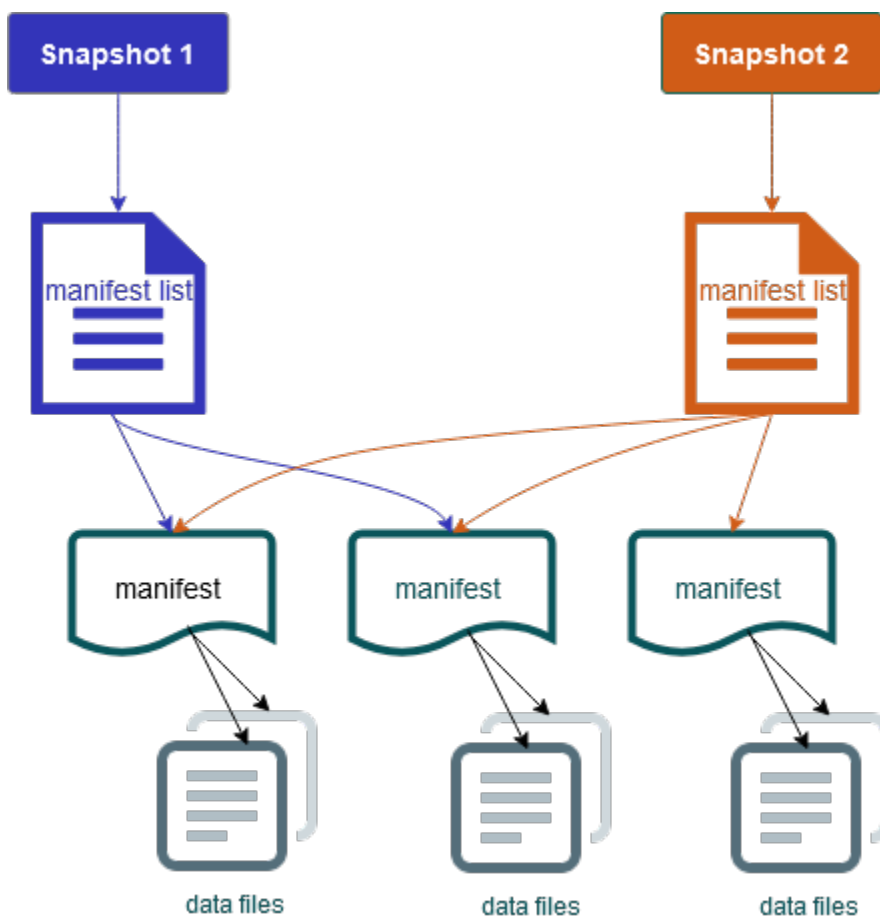
- [Funzionamento di Iceberg](#)
- [Creazione di un cluster con Iceberg installato](#)
- [Considerazioni e limitazioni per l'utilizzo di Iceberg su Amazon EMR](#)
- [Cronologia dei rilasci di Iceberg](#)

Funzionamento di Iceberg

Iceberg tiene traccia dei singoli file di dati in una tabella anziché nelle directory. In questo modo, gli scrittori possono creare file di dati sul posto (i file non vengono spostati o modificati). Inoltre, possono aggiungere file alla tabella solo in un commit esplicito. Lo stato della tabella viene mantenuto nei file di metadati. Tutte le modifiche allo stato della tabella creano un nuovo file di metadati e sostituiscono atomicamente i metadati precedenti. Il file di metadati della tabella tiene traccia dello schema della tabella, della configurazione del partizionamento e di altre proprietà.

Include anche istantanee del contenuto della tabella. Ogni snapshot è un insieme completo di file di dati nella tabella in un determinato momento. Gli snapshot sono elencati nel file di metadati, ma i file in uno snapshot vengono archiviati in file manifesti separati. Le transizioni atomiche dal file di metadati di una tabella al successivo forniscono l'isolamento degli snapshot. I lettori utilizzano lo snapshot corrente quando hanno caricato i metadati della tabella. I lettori non sono interessati dalle modifiche finché non eseguono l'aggiornamento e raccolgono una nuova posizione dei metadati. I file di dati negli snapshot vengono archiviati in uno o più file manifesti che contengono una riga per ciascun file di dati nella tabella, i relativi dati di partizione e i relativi parametri. Uno snapshot è l'unione di tutti i file nei suoi manifesti. I file manifesti possono anche essere condivisi tra le istantanee per evitare di riscrivere i metadati che cambiano lentamente.

Diagramma degli snapshot di Iceberg



Iceberg offre le seguenti funzionalità:

- Supporta le transazioni ACID e i viaggi nel tempo nel tuo data lake Amazon S3.
- I tentativi di commit beneficiano dei vantaggi prestazionali di [concorrenza ottimistica](#).
- La risoluzione dei conflitti a livello di file comporta un'elevata concorrenza.

- Con le statistiche minima e massima per colonna nei metadati puoi ignorare i file, aumentando le prestazioni per le query selettive.
- Puoi organizzare le tabelle in layout di partizioni flessibili con l'evoluzione delle partizioni che consente di aggiornare gli schemi di partizioni. Le query e i volumi di dati possono quindi cambiare senza fare affidamento su directory fisiche.
- Supporta l'applicazione e l'[evoluzione dello schema](#).
- Le tabelle Iceberg fungono da dissipatori idempotenti e fonti riproducibili. Ciò consente lo streaming e il supporto batch con pipeline di elaborazione singola. I dissipatori idempotenti tengono traccia delle operazioni di scrittura che hanno avuto successo in passato. Pertanto, il dissipatore può richiedere nuovamente i dati in caso di errore ed eliminarli se sono stati inviati più volte.
- Cronologia e derivazione sono visibili, comprendendo l'evoluzione della tabella, la cronologia delle operazioni e le statistiche per ogni commit.
- Migrazione da un set di dati esistente con una scelta di formato di dati (Parquet, ORC, Avro) e motore di analisi (Spark, Trino, PrestoDB, Flink, Hive).

Creazione di un cluster con Iceberg installato

Questa sezione include informazioni sull'utilizzo di Iceberg con Spark, Trino, Flink e Hive.

Utilizza un cluster Iceberg con Spark

A partire da Amazon EMR versione 6.5.0, puoi utilizzare Iceberg con il cluster Spark senza dover includere operazioni di bootstrap. Per Amazon EMR versione 6.4.0 e precedenti, puoi utilizzare un'operazione di bootstrap per preinstallare tutte le dipendenze necessarie.

In questo tutorial, usa AWS CLI per lavorare con Iceberg su un cluster Amazon EMR Spark. Per utilizzare la console per creare un cluster con Iceberg installato, segui la procedura illustrata in [Creazione di un data lake Apache Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue](#).

Creazione di un cluster Iceberg

Puoi creare un cluster con Iceberg installato utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. In questo tutorial, usa la AWS CLI per lavorare con Iceberg su un cluster Amazon EMR. Per utilizzare la console per creare un cluster con Iceberg installato, segui la procedura illustrata in [Creazione di un data lake Apache Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue](#).

Per utilizzare Iceberg su Amazon EMR con la AWS CLI, crea prima un cluster completando la seguente procedura. Per informazioni su come specificare la classificazione Iceberg utilizzando la AWS CLI, consulta [Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster](#) o [Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster](#).

1. Crea un file `configurations.json` con i seguenti contenuti:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

2. Quindi, crea un cluster con la seguente configurazione. Sostituisci il percorso del bucket Amazon S3 di esempio e l'ID della sottorete con i tuoi valori.

```
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
```

In alternativa, puoi creare un cluster Amazon EMR che include l'applicazione Spark e includere il file `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` come dipendenza JAR in un processo Spark. Per ulteriori informazioni, consulta [Invio di applicazioni](#).

Per includere il jar come dipendenza in un processo Spark, aggiungi la seguente proprietà di configurazione all'applicazione Spark:

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

Per ulteriori informazioni sulle dipendenze dei processi Spark, consulta [Dependency Management](#) (Gestione delle dipendenze) nel documento di Apache Spark [Running Spark on Kubernetes](#) (Esecuzione di Spark su Kubernetes).

Inizializzazione di una sessione Spark per Iceberg

Negli esempi seguenti viene illustrato come avviare la shell interattiva Spark, utilizzare Spark submit o utilizzare Amazon EMR Notebooks per lavorare con Iceberg su Amazon EMR.

spark-shell

1. Connessione al nodo master tramite SSH Per ulteriori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
2. Immettere il seguente comando per avviare la shell Spark. Per utilizzare la shell PySpark, sostituire `spark-shell` con `pyspark`.

```
spark-shell \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

spark-submit

1. Connessione al nodo master tramite SSH Per ulteriori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
2. Immettere il seguente comando per avviare la sessione Spark per Iceberg.

```
spark-submit \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

EMR Studio notebooks

Per inizializzare una sessione Spark utilizzando i notebook EMR Studio, configurare la sessione Spark utilizzando il comando magico `%%configure` sul tuo notebook Amazon EMR, come

nell'esempio seguente. Per ulteriori informazioni, consulta [Utilizzo di EMR Notebooks magics](#) nella Guida alla gestione di Amazon EMR.

```
%%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/"
}
}
```

Scrittura su una tabella Iceberg

Nell'esempio seguente viene illustrato come creare un DataFrame e scriverlo come set di dati Iceberg. In questa sezione, gli esempi illustrano l'utilizzo di set di dati con la Spark shell connessi al nodo master utilizzando SSH come utente hadoop predefinito.

Note

Per incollare gli esempi di codice nella shell Spark, digitare `:paste` al prompt, incollare l'esempio e premere CTRL+D.

PySpark

Spark include anche una shell basata su Python, `pyspark`, che puoi utilizzare per realizzare prototipi di programmi Spark scritti in Python. Richiama `pyspark` sul nodo principale.

```
## 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()
```

Lettura da una tabella Iceberg

PySpark

```
df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Scala

```
val df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Spark SQL

```
SELECT * from dev.db.iceberg_table LIMIT 10
```

Configurare le proprietà Spark per utilizzare il Catalogo dati di Glue di AWS come metastore delle tabelle Iceberg

Per utilizzare AWS Glue Catalog come metastore per le tabelle Iceberg, imposta le proprietà di configurazione Spark come di seguito:

```
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
```

Utilizza un cluster Iceberg con Trino

A partire da Amazon EMR versione 6.6.0, puoi utilizzare Iceberg con il cluster Trino.

In questo tutorial, usa la AWS CLI per lavorare con Iceberg su un cluster Amazon EMR Trino. Per utilizzare la console per creare un cluster con Iceberg installato, segui la procedura illustrata in [Creazione di un data lake Apache Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue](#).

Creazione di un cluster Iceberg

Per utilizzare Iceberg su Amazon EMR con la AWS CLI, crea prima un cluster completando la seguente procedura. Per informazioni su come specificare la classificazione Iceberg utilizzando la AWS CLI, consulta [Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster](#) o [Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster](#).

1. Crea un file `iceberg.properties` e imposta un valore per il catalogo scelto. Ad esempio, se desideri utilizzare il metastore Hive come catalogo, il file deve avere il seguente contenuto.

```
connector.name=iceberg
```

```
hive.metastore.uri=thrift://localhost:9083
```

Se desideri utilizzare il Catalogo dati AWS Glue come archivio, il file deve avere il seguente contenuto.

```
connector.name=iceberg
iceberg.catalog.type=glue
```

2. Crea un'operazione di bootstrap che copi `iceberg.properties` da Amazon S3 su `/etc/trino/conf/catalog/iceberg.properties`, come nell'esempio seguente. Per ulteriori informazioni sulle operazioni di bootstrap, consulta [Crea operazioni bootstrap per installare software aggiuntivi](#).

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/iceberg.properties
```

3. Crea un cluster con la seguente configurazione, sostituendo il percorso dello script delle operazioni bootstrap e il nome della chiave di esempio con i tuoi valori.

```
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>
```

Inizializzazione di una sessione Trino per Iceberg

Per inizializzare una sessione Trino, esegui il comando riportato di seguito.

```
trino-cli --catalog iceberg
```

Scrittura su una tabella Iceberg

Crea e scrivi nella tabella con i comando SQL seguenti.


```
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');
```

Leggi da una tabella Iceberg

Per leggere dalla tabella Iceberg, esegui il comando seguente.

```
trino> SELECT * from default.iceberg_table;
```

Utilizzo di un cluster Iceberg con Flink

A partire dalla versione 6.9.0 di Amazon EMR, puoi utilizzare Iceberg con un cluster Flink senza i passaggi di configurazione necessari quando si utilizza l'integrazione open source di Iceberg per Flink.

Creazione di un cluster Iceberg

Puoi creare un cluster con Iceberg installato utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. In questo tutorial, usa la AWS CLI per lavorare con Iceberg su un cluster Amazon EMR. Per utilizzare la console per creare un cluster con Iceberg installato, segui la procedura illustrata in [Creazione di un data lake Apache Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue](#).

Per utilizzare Iceberg su Amazon EMR con la AWS CLI, crea prima un cluster completando la seguente procedura. Per informazioni su come specificare la classificazione Iceberg utilizzando la AWS CLI, consulta [Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster](#) o [Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster](#). Crea un file denominato `configurations.json` con i seguenti contenuti:

```
[{
```

```
"Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Quindi, crea un cluster con la seguente configurazione, sostituendo il percorso del bucket Amazon S3 di esempio e l'ID della sottorete con i tuoi valori:

```
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
```

In alternativa, puoi creare un cluster Amazon EMR 6.9.0 che include l'applicazione Flink e utilizza il file `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` come dipendenza JAR in un processo Flink.

Utilizzo del client Flink SQL

Lo script SQL Client si trova in `/usr/lib/flink/bin`. Puoi eseguire lo script con il comando seguente:

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

Questa operazione avvia una shell (interprete di comandi) Flink SQL.

Esempi di Flink

Creazione di una tabella Iceberg

Flink SQL

```
CREATE CATALOG glue_catalog WITH (
  'type'='iceberg',
  'warehouse'='<WAREHOUSE>',
  'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
```

```

    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO',
    'lock-impl'='org.apache.iceberg.aws.glue.DynamoLockManager',
    '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);

```

Tabella 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);");

```

Scrittura su una tabella Iceberg

Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

Tabella API

```
tEnv.executeSql(
    "INSERT INTO `glue_catalog`.`"
    + db
    + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

API del flusso di dati

```
final StreamExecutionEnvironment env =
    StreamExecutionEnvironment.getExecutionEnvironment();

final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

GenericRowData rowData1 = new GenericRowData(2);
rowData1.setField(0, 1L);
rowData1.setField(1, StringData.fromString("a"));

DataStream<RowData> input = env.fromElements(rowData1);

Map<String, String> props = new HashMap<>();
props.put("type", "iceberg");
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatlogLoader =
    CatalogLoader.custom(
        "glue",
        props,
        new Configuration(),
        "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
    TableLoader.fromCatalog(glueCatlogLoader, TableIdentifier.of(db, "sample"));

DataStreamSink<Void> dataStreamSink =
```

```
FlinkSink.forRowData(input).tableLoader(tableLoader).append();

env.execute("Datastream Write");
```

Letture da una tabella Iceberg

Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

Tabella API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

API del flusso di dati

```
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");
```

Utilizzo del catalogo Hive

Assicurati che le dipendenze di Flink e Hive siano risolte come descritto in [Configurazione di Flink con Hive Metastore e Catalogo Glue](#).

Esecuzione di un processo Flink

Un modo per inviare un processo a Flink consiste nell'utilizzare una sessione YARN per il processo Flink. Puoi eseguire l'avvio tramite il comando seguente:

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

Utilizzo di un cluster Iceberg con Hive

Con le versioni 6.9.0 e successive di Amazon EMR, puoi utilizzare Iceberg con un cluster Hive senza dover eseguire i passaggi di configurazione necessari per l'integrazione open source di Iceberg per Hive. Per Amazon EMR versione 6.8.0 e precedenti, puoi utilizzare un'operazione di bootstrap per installare `jar iceberg-hive-runtime` al fine di configurare Hive per il supporto di Iceberg.

Amazon EMR 6.9.0 include tutte le funzionalità per [l'integrazione di Hive 3.1.3 con Iceberg 0.14.1](#) e include inoltre funzionalità aggiuntive di Amazon EMR come la selezione automatica dei motori di esecuzione supportati in fase di runtime (Amazon EMR su EKS 6.9.0).

Creazione di un cluster Iceberg

Puoi creare un cluster con Iceberg installato utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. In questo tutorial, usa la AWS CLI per lavorare con Iceberg su un cluster Amazon EMR. Per utilizzare la console al fine di creare un cluster con Iceberg installato, segui la procedura illustrata in [Creazione di un data lake Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue](#).

Per utilizzare Iceberg su Amazon EMR con la AWS CLI, crea prima un cluster tramite la seguente procedura. Per informazioni su come specificare la classificazione Iceberg utilizzando la AWS CLI o l'SDK Java, consulta [Fornitura di una configurazione utilizzando AWS CLI quando si crea un cluster](#) o [Fornitura di una configurazione utilizzando l'SDK Java per la creazione di un cluster](#). Crea un file denominato `configurations.json` con i seguenti contenuti:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Quindi, crea un cluster con la seguente configurazione, sostituendo il percorso del bucket Amazon S3 di esempio e l'ID della sottorete con i tuoi valori:

```
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
```

Il cluster Hive Iceberg è in grado di eseguire le seguenti operazioni:

- Caricare il jar di runtime di Iceberg Hive in Hive e abilitare la configurazione relativa a Iceberg per il motore Hive.
- Abilitare la selezione dinamica del motore di esecuzione di Amazon EMR Hive per impedire agli utenti di impostare il motore di esecuzione supportato per la compatibilità con Iceberg.

Note

I cluster Hive Iceberg attualmente non supportano il AWS Glue Data Catalog. Il catalogo Iceberg predefinito è HiveCatalog, che corrisponde al metastore configurato per l'ambiente Hive. Per ulteriori informazioni sulla gestione dei cataloghi, consulta [Catalog Management](#) (Gestione dei cataloghi) nella [documentazione di Apache Hive](#).

Supporto funzionalità

Amazon EMR 6.9.0 supporta Hive 3.1.3 e Iceberg 0.14.1. Il supporto delle funzionalità è limitato alle funzioni compatibili con Iceberg per Hive 3.1.2 e 3.1.3. Sono supportati i seguenti comandi:

- Con le versioni da 6.9.0 a 6.12.x di Amazon EMR, devi includere il jar `libfb303` nella directory `Hive auxlib`. Utilizza il comando seguente per includerlo:

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/libfb303.jar
```

Con le versioni 6.13 e successive di Amazon EMR, il jar `libfb303` viene automaticamente collegato simbolicamente alla directory `Hive auxlib`.

- Creazione di una tabella
 - Tabella non partizionata: è possibile creare tabelle esterne in Hive fornendo il gestore di archiviazione come segue:

```
CREATE EXTERNAL TABLE x (i int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- Tabella partizionata: è possibile creare tabelle partizionate esterne in Hive nel modo seguente:

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

Note

Il formato di file `STORED AS` per `ORC/AVRO/PARQUET` non è supportato in Hive 3. L'unica opzione predefinita è `Parquet`.

- Eliminazione di una tabella: il comando `DROP TABLE` viene utilizzato per eliminare tabelle, come nell'esempio seguente:

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- Lettura di una tabella: le istruzioni `SELECT` possono essere utilizzate per leggere le tabelle Iceberg in Hive, come nell'esempio seguente. I motori di esecuzione supportati sono `MR` e `Tez`.

```
SELECT * FROM table_name
```

Per informazioni sulla sintassi di selezione di Hive, consulta [LanguageManual Select](#). Per informazioni sulle istruzioni selezionate con tabelle Iceberg in Hive, consulta [Apache Iceberg Select](#).

- Inserimento in una tabella: l'istruzione `INSERT INTO` di HiveQL funziona solo su tabelle Iceberg con supporto per il motore di esecuzione Map Reduce. Gli utenti di Amazon EMR non devono impostare esplicitamente il motore di esecuzione dal momento che Amazon EMR Hive lo seleziona automaticamente in fase di runtime.
- Inserimento di una singola tabella. Esempio:

```
INSERT INTO table_name VALUES ('a', 1);
INSERT INTO table_name SELECT...
```

- Inserimento di più tabelle: sono supportati gli inserimenti multi-tabella non atomici nelle istruzioni. Esempio:

```
FROM source
INSERT INTO table_1 SELECT a, b
INSERT INTO table_2 SELECT c,d;
```

Considerazioni e limitazioni per l'utilizzo di Iceberg su Amazon EMR

Questa sezione include considerazioni e limitazioni per l'utilizzo di Iceberg con Spark, Trino, Flink e Hive.

Considerazioni sull'utilizzo di Iceberg con Spark

- Per impostazione predefinita Amazon EMR 6.5.0 non supporta Iceberg in esecuzione su Amazon EMR su EKS. È disponibile un'immagine personalizzata di Amazon EMR 6.5.0 che ti consente di passare `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` come parametro `spark-submit` per creare tabelle Iceberg in Amazon EMR su EKS. Per ulteriori informazioni, consulta [Invio di un carico di lavoro Spark in Amazon EMR utilizzando un'immagine personalizzata](#) nella Guida per gli sviluppatori di Amazon EMR su EKS. Puoi anche contattare AWS Support per ricevere assistenza. A partire da Amazon EMR 6.6.0, Iceberg è supportato in Amazon EMR su EKS.
- Quando utilizzi Glue di AWS come catalogo per Iceberg, assicurati che il database in cui si sta creando una tabella esista in Glue AWS. Se utilizzi servizi come AWS Lake Formation e non riesci a caricare il catalogo, assicurati di avere accesso adeguato al servizio per eseguire il comando.

Considerazioni sull'utilizzo di Iceberg con Trino

- Amazon EMR 6.5 non offre il supporto del Catalogo Trino Iceberg per Iceberg in modo nativo. Trino necessita di Iceberg v0.11, quindi ti consigliamo di avviare un cluster Amazon EMR diverso per Trino dal cluster Spark e di includere Iceberg v0.11 su tale cluster.
- Quando utilizzi AWS Glue come catalogo per Iceberg, assicurati che il database in cui si sta creando una tabella sia presente in AWS Glue. Se utilizzi servizi come AWS Lake Formation e non riesci a caricare il catalogo, assicurati di avere accesso adeguato al servizio per eseguire il comando.

Considerazioni sull'utilizzo di Iceberg con Flink

Quando utilizzi Glue di AWS come catalogo per Iceberg, assicurati che il database in cui si sta creando una tabella esista in Glue AWS. Se utilizzi servizi come AWS Lake Formation e non riesci a caricare il catalogo, assicurati di avere accesso adeguato al servizio per eseguire il comando.

Considerazioni sull'utilizzo di Iceberg con Hive

- Iceberg supporta i seguenti tipi di query:
 - Create table (Crea tabella)
 - Drop table (Elimina tabella)
 - Insert into table (Inserisci in tabella)
 - Read table (Leggi tabella)
- Solo il motore di esecuzione MR (MapReduce) è supportato per le operazioni DML (Data Manipulation Language) e MR è obsoleto in Hive 3.1.3.
- Il Catalogo dati di AWS Glue Data Catalog non è attualmente supportato per Iceberg con Hive.
- La gestione degli errori non è sufficientemente robusta. In caso di configurazione errata, gli inserimenti nelle query potrebbero essere completati correttamente. Tuttavia, il mancato aggiornamento dei metadati può causare la perdita di dati.

Cronologia dei rilasci di Iceberg

Nella seguente tabella sono elencati la versione di Iceberg inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Iceberg

Etichetta di rilascio di Amazon EMR	Versione di Iceberg	Componenti installati con Iceberg
emr-6.14.0	1.3.1-amzn-0	Not available.
emr-6.13.0	1.3.0-amzn-1	Not available.
emr-6.12.0	1.3.0-amzn-0	Not available.
emr-6.11.1	1.2.0-amzn-0	Not available.
emr-6.11.0	1.2.0-amzn-0	Not available.
emr-6.10.1	1.1.0-amzn-0	Not available.
emr-6.10.0	1.1.0-amzn-0	Not available.
emr-6.9.1	0.14.1-amzn-0	Not available.
emr-6.9.0	0.14.1-amzn-0	Not available.
emr-6.8.1	0.14.0-amzn-0	Not available.
emr-6.8.0	0.14.0-amzn-0	Not available.
emr-6.7.0	0.13.1-amzn-0	Not available.
emr-6.6.0	0.13.1	Not available.
emr-6.5.0	0.12.0	Not available.

Note sul rilascio di Iceberg per versione

- [Amazon EMR 6.9.0 - Note sul rilascio per Iceberg](#)

Amazon EMR 6.9.0 - Note sul rilascio per Iceberg

Amazon EMR 6.9.0 - Modifiche a Iceberg

Tipo	Descrizione
Caratteristica	Integrazione di Amazon EMR Flink con Iceberg.
Caratteristica	Integrazione di Amazon EMR Hive con Iceberg.
Caratteristica	È stato aggiunto il supporto per memorizzare nella cache i file di metadati Iceberg su Amazon FSx per Lustre per migliorare i tempi di pianificazione delle query.
Backport	PR 5050 : Flink 1.15: è stato aggiunto il supporto per le opzioni di scrittura nei commenti SQL di inserimento inline.
Backport	PR 5282 : AWS: sono stati corretti gli errori di ripetizione PUT aprendo nuovi flussi di file di dati.
Backport	PR 5318 : Flink 1.15: è stato colmato il divario tra FlinkSource e IcebergSource (FLIP-27) ed è stata aggiunta una configurazione facoltativa per utilizzare la sorgente FLIP-27 in Flink SQL.
Backport	PR 5344 : Flink 1.14: è stato colmato il divario tra FlinkSource e IcebergSource (FLIP-27) ed è stata aggiunta una configurazione facoltativa per utilizzare la sorgente FLIP-27 in Flink SQL.

Tipo	Descrizione
Backport	PR 5393 : Flink 1.14, 1.15: non è più necessario convertire i parametri di Iceberg MetricContext in Flink nel lettore sorgente FLIP-27.
Backport	PR 5401 : Flink 1.14, 1.15: il gruppo IcebergSourceReader in PR #5393 per i parametri del lettore sorgente FLIP-27 era assente.
Backport	PR 5679 : Spark 3.2, 3.3: è stata corretta la propagazione del supporto dei valori Null per il nodo MergeRows.
Backport	PR 5860 : Spark 3.3: è stato corretto l'errore QueryFailure durante l'esecuzione di RewriteManifestProcedures sulle tabelle partizionate Date.
Backport	PR 5880 : Spark 3.3: è stato corretto il supporto dei valori Null nelle proiezioni merge-on-read.
Backport	PR 5917 : Spark 3.2: è stato corretto il supporto dei valori Null nelle proiezioni merge-on-read.

Notebook Jupyter su Amazon EMR

[Jupyter Notebook](#) è un'applicazione Web open source che gli utenti possono utilizzare per creare e condividere documenti contenenti codice live, equazioni, visualizzazioni e testo introduttivo. Amazon EMR offre tre opzioni per lavorare con i notebook Jupyter:

Argomenti

- [EMR Studio](#)
- [Notebook Amazon EMR basato su notebook Jupyter](#)
- [JupyterHub](#)

EMR Studio

Amazon EMR Studio è un ambiente di sviluppo integrato (IDE) basato sul Web per [notebook Jupyter](#) completamente gestiti che vengono eseguiti su cluster Amazon EMR. Puoi configurare un EMR Studio per il tuo team per sviluppare, visualizzare ed eseguire il debug di applicazioni scritte in R, Python, Scala e PySpark.

Ti consigliamo di utilizzare EMR Studio quando utilizzi i notebook Jupyter su Amazon EMR. Per ulteriori informazioni, consulta l'argomento [EMR Studio](#) della Guida alla gestione di Amazon EMR.

Notebook Amazon EMR basato su notebook Jupyter

EMR Notebooks è un ambiente [Jupyter Notebook](#) integrato nella console di Amazon EMR che consente di creare rapidamente notebook Jupyter, collegarli a cluster Spark e aprire l'editor Jupyter Notebook nella console per eseguire in remoto le query e il codice. Un notebook EMR viene salvato in Amazon S3 indipendentemente dai cluster per un'archiviazione duratura, un accesso rapido e un'elevata flessibilità. È possibile disporre di più notebook, collegare più notebook a un singolo cluster e riutilizzare un notebook in cluster diversi.

Per ulteriori informazioni, consulta [Notebook EMR](#) nella Guida alla gestione di Amazon EMR.

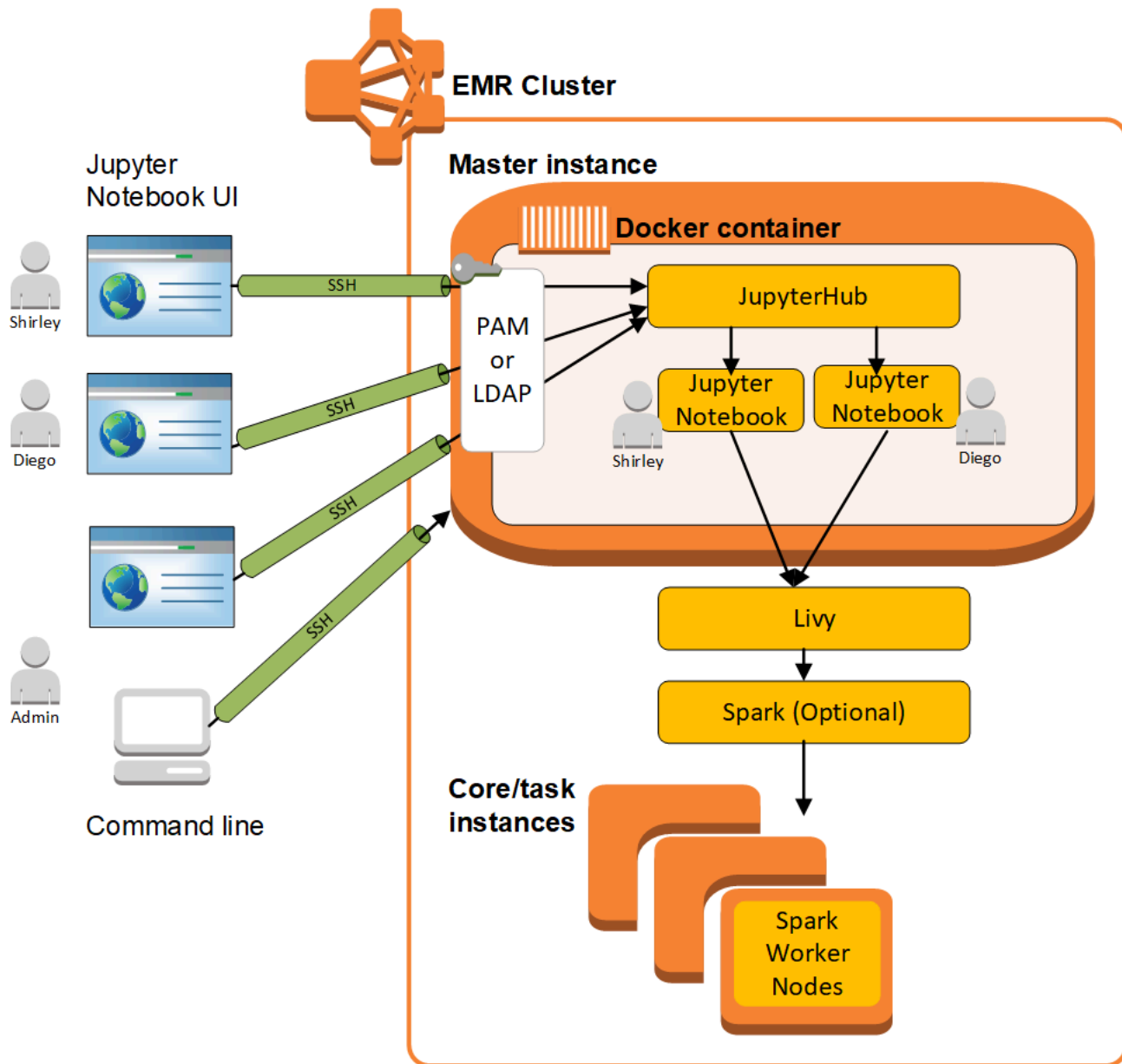
JupyterHub

[Jupyter Notebook](#) è un'applicazione Web open source che puoi utilizzare per creare e condividere documenti contenenti il codice live, equazioni, visualizzazioni e testo introduttivo. [JupyterHub](#)

consente di ospitare più istanze di un server del notebook Jupyter a un solo utente. Quando crei un cluster con JupyterHub, Amazon EMR crea un container Docker sul nodo master del cluster. JupyterHub, tutti i componenti necessari per Jupyter e [Sparkmagic](#) vengono eseguiti all'interno del container.

Sparkmagic è una libreria di kernel che consente ai notebook Jupyter di interagire con [Apache Spark](#) in esecuzione su Amazon EMR tramite [Apache Livy](#), che è un server REST per Spark. Spark e Apache Livy vengono installati automaticamente al momento della creazione di un cluster con JupyterHub. Il kernel Python 3 predefinito per Jupyter è disponibile insieme ai kernel PySpark 3, PySpark e Spark disponibili con Sparkmagic. Puoi utilizzare questi kernel per eseguire codice Spark ad-hoc e query SQL interattive utilizzando Python e Scala. Puoi installare kernel aggiuntivi all'interno del container Docker manualmente. Per ulteriori informazioni, consulta [Installazione di kernel e librerie aggiuntive](#).

Il seguente diagramma mostra i componenti di JupyterHub su Amazon EMR con metodi di autenticazione corrispondenti per gli utenti del notebook e l'amministratore. Per ulteriori informazioni, consulta [Aggiunta di utenti e amministratori Jupyter Notebook](#).



La seguente tabella indica la versione di JupyterHub inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con JupyterHub.

Per la versione dei componenti installati con JupyterHub in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di JupyterHub per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con JupyterHub
emr-6.14.0	JupyterHub 1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

La seguente tabella indica la versione di JupyterHub inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con JupyterHub.

Per la versione dei componenti installati con JupyterHub in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di JupyterHub per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con JupyterHub
emr-5.36.1	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,

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con JupyterHub
		hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Il kernel Python 3 incluso con JupyterHub su Amazon EMR è 3.6.4.

Le librerie installate all'interno del container `jupyterhub` possono variare tra versioni del rilascio Amazon EMR e versioni dell'AMI Amazon EC2.

Per elencare le librerie installate utilizzando **conda**

- Eseguire il seguente comando sulla riga di comando del nodo master:

```
sudo docker exec jupyterhub bash -c "conda list"
```

Per elencare le librerie installate utilizzando **pip**

- Eseguire il seguente comando sulla riga di comando del nodo master:

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Argomenti

- [Creazione di un cluster con JupyterHub](#)
- [Considerazioni sull'utilizzo di JupyterHub su Amazon EMR](#)
- [Configurazione di JupyterHub](#)
- [Configurazione della persistenza per i notebook in Amazon S3](#)
- [Connessione al nodo master e ai server notebook](#)
- [Configurazione e amministrazione di JupyterHub](#)
- [Aggiunta di utenti e amministratori Jupyter Notebook](#)
- [Installazione di kernel e librerie aggiuntive](#)

- [Cronologia dei rilasci di JupyterHub](#)

Creazione di un cluster con JupyterHub

È possibile creare un cluster Amazon EMR con JupyterHub utilizzando la AWS Management Console, la AWS Command Line Interface o l'API di Amazon EMR. Assicurati che il cluster non venga creato con l'opzione per terminare automaticamente dopo il completamento delle fasi (opzione `--auto-terminate` in AWS CLI). Inoltre, assicurati che gli amministratori e gli utenti di notebook siano in grado di accedere alla coppia di chiavi utilizzata durante la creazione del cluster. Per ulteriori informazioni, consulta la sezione relativa all'[Utilizzo di una coppia di chiavi per le credenziali SSH](#) nella Guida per la gestione di Amazon EMR.

Creazione di un cluster con JupyterHub utilizzando la console

Utilizza la seguente procedura per creare un cluster con JupyterHub installato utilizzando Advanced Options (Opzioni avanzate) nella console di Amazon EMR.

Creazione di un cluster Amazon EMR con JupyterHub installato utilizzando la console di Amazon EMR

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Software Configuration (Configurazione del software):
 - In Versione, seleziona emr-5.36.1 e scegli JupyterHub.
 - Se si utilizza Spark, per utilizzare AWS Glue Data Catalog come metastore per Spark SQL occorre selezionare Use for Spark table metadata (Utilizza per i metadati delle tabelle Spark). Per ulteriori informazioni, consulta [Utilizzare AWS Glue Data Catalog come metastore Spark SQL](#).
 - Per Edit software settings (Modifica impostazioni software), scegliere Enter configuration (Immetti configurazione) e specificare i valori di configurazione o scegliere Load JSON from S3 (Carica JSON da S3) e specificare un file di configurazione JSON. Per ulteriori informazioni, consulta [Configurazione di JupyterHub](#).
4. In Add steps (optional) (Aggiungi fasi (opzionale)) configurare le fasi da eseguire quando il cluster viene creato, accertarsi che Auto-terminate cluster after the last step is completed (Una

- volta completata la fase, termina automaticamente il cluster) non sia selezionata e scegliere Next (Successivo).
- Scegliere le opzioni Hardware Configuration (Configurazione hardware), Next (Successivo). Per ulteriori informazioni, consulta la sezione relativa alla [Configurazione di hardware e reti cluster](#) nella Guida per la gestione di Amazon EMR.
 - Scegliere opzioni per General Cluster Settings (Impostazioni generali del cluster), Next (Successivo).
 - Scegliere Security Options (Opzioni di protezione), specificando un coppia di chiavi e scegliere Create Cluster (Crea cluster).

Creazione di un cluster con JupyterHub utilizzando la AWS CLI

Per avviare un cluster con JupyterHub, utilizza il comando `aws emr create-cluster` e per l'opzione `--applications` specifica `Name=JupyterHub`. Nell'esempio seguente viene avviato un cluster JupyterHub su Amazon EMR con due istanze EC2 (un'istanza master e un'istanza principale). Inoltre, il debug è abilitato, con i log archiviati nel percorso Amazon S3 come specificato da `--log-uri`. La coppia di chiavi specificata fornisce l'accesso a istanze Amazon EC2 nel cluster.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.1 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes  
KeyName=MyKeyPair
```

Considerazioni sull'utilizzo di JupyterHub su Amazon EMR

Considera le informazioni seguenti durante l'utilizzo di JupyterHub su Amazon EMR.

- **Warning**

I notebook e i file utente vengono salvati nel file system sul nodo master. Si tratta di uno storage temporaneo che non viene mantenuto attraverso la terminazione del cluster. Quando un cluster viene terminato, questi dati vengono persi se non si esegue un backup. Ti consigliamo di pianificare l'esecuzione di backup regolari utilizzando processi cron o un altro mezzo idoneo per l'applicazione.

Inoltre, modifiche di configurazione apportate all'interno del container potrebbero non persistere in caso di riavvio del container. Ti consigliamo di creare uno script o automatizzare la configurazione del container in modo da riprodurre le personalizzazioni più facilmente.

- L'autenticazione Kerberos che è stata configurata utilizzando una configurazione di sicurezza Amazon EMR non è supportata.
- [OAuthenticator](#) non è supportato.

Configurazione di JupyterHub

Puoi personalizzare la configurazione di JupyterHub su Amazon EMR e singoli notebook utente effettuando la connessione al nodo master del cluster e modificando i file di configurazione. Dopo aver modificato i valori, riavvia il container jupyterhub.

Modifica le proprietà nei seguenti file per configurare JupyterHub e notebook Jupyter singoli:

- `jupyterhub_config.py`: per impostazione predefinita, questo file viene salvato nella directory `/etc/jupyter/conf/` sul nodo master. Per ulteriori informazioni, consulta la sezione relativa alle [Nozioni di base sulla configurazione](#) nella documentazione di JupyterHub.
- `jupyter_notebook_config.py`: questo file viene salvato nella directory `/etc/jupyter/` per impostazione predefinita e copiato nel container jupyterhub come predefinito. Per ulteriori informazioni, consulta la sezione relativa al [file di configurazione e alle opzioni della riga di comando](#) nella documentazione di Jupyter Notebook.

Puoi anche utilizzare la classificazione di configurazione `jupyter-sparkmagic-conf` per personalizzare Sparkmagic, che consente di aggiornare i valori nel file `config.json` per Sparkmagic. Per ulteriori informazioni sulle impostazioni disponibili, consulta [example_config.json su](#)

[GitHub](#). Per ulteriori informazioni sull'utilizzo delle classificazioni di configurazione con le applicazioni in Amazon EMR, consulta [Configurazione delle applicazioni](#).

Nell'esempio seguente viene avviato un cluster utilizzando AWS CLI, facendo riferimento al file `MyJupyterConfig.json` per le impostazioni di classificazione di configurazione di Sparkmagic.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
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
```

Di seguito sono mostrati i contenuti di esempio di `MyJupyterConfig.json`:

```
[
  {
    "Classification": "jupyter-sparkmagic-conf",
    "Properties": {
      "kernel_python_credentials" : "{\"username\": \"diego\", \"base64_password\": \"mypass\", \"url\": \"http://localhost:8998\", \"auth\": \"None\"}"
    }
  }
]
```

Note

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Configurazione della persistenza per i notebook in Amazon S3

Puoi configurare un cluster JupyterHub in Amazon EMR in modo che i notebook salvati da un utente persistano in Amazon S3, al di fuori dell'archiviazione temporanea sulle istanze EC2 del cluster.

La persistenza di Amazon S3 viene specificata utilizzando la classificazione di configurazione `jupyter-s3-conf` al momento della creazione di un cluster. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Oltre ad abilitare la persistenza di Amazon S3 utilizzando la proprietà `s3.persistence.enabled`, è possibile specificare un bucket in Amazon S3 in cui i notebook vengono salvati utilizzando la proprietà `s3.persistence.bucket`. I notebook per ciascun utente vengono salvati in una cartella `jupyter/jupyterhub-user-name` nel bucket specificato. Il bucket deve esistere già in Amazon S3 e il ruolo per il profilo dell'istanza EC2 specificato al momento della creazione del cluster deve disporre delle autorizzazioni per il bucket (per impostazione predefinita, il ruolo è `EMR_EC2_DefaultRole`). Per ulteriori informazioni, consulta [Configurazione di ruoli IAM per le autorizzazioni Amazon EMR ai servizi AWS](#).

Quando avvii un nuovo cluster utilizzando le stesse proprietà di classificazione di configurazione, gli utenti possono aprire i notebook con i contenuti dalla posizione salvata.

Tieni presente che l'importazione del file sotto forma di moduli in un notebook quando Amazon S3 è abilitato comporterà il caricamento dei file su Amazon S3. Quando importi file senza abilitare la persistenza di Amazon S3, questi vengono caricati nel container JupyterHub.

L'esempio seguente abilita la persistenza di Amazon S3. I notebook salvati dagli utenti vengono salvati nella cartella `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` per ciascun utente, dove `jupyterhub-user-name` è un nome utente, ad esempio `diego`.

```
[
  {
    "Classification": "jupyter-s3-conf",
    "Properties": {
      "s3.persistence.enabled": "true",
      "s3.persistence.bucket": "MyJupyterBackups"
    }
  }
]
```

Connessione al nodo master e ai server notebook

Gli amministratori JupyterHub e gli utenti di notebook devono connettersi al nodo master del cluster utilizzando un tunnel SSH ed effettuando la connessione alle interfacce Web servite da JupyterHub sul nodo master. Per ulteriori informazioni sulla configurazione di un tunnel SSH e l'utilizzo del tunnel per effettuare il proxy delle connessioni Web, consulta [Connessione al cluster](#) nella Guida per la gestione di Amazon EMR.

Per impostazione predefinita, JupyterHub su Amazon EMR è disponibile tramite la porta 9443 sul nodo master. Il proxy JupyterHub interno serve anche istanze notebook tramite la porta 9443. L'accesso alle interfacce Web JupyterHub e Jupyter è possibile tramite un URL con il seguente modello:

```
https://MasterNodeDNS:9443
```

Puoi specificare una porta differente utilizzando la proprietà `c.JupyterHub.port` nel file `jupyterhub_config.py`. Per ulteriori informazioni, consulta la sezione relativa alle [Networking basics \(Nozioni di base sulle reti\)](#) nella documentazione di JupyterHub.

Per impostazione predefinita, JupyterHub su Amazon EMR utilizza un certificato autofirmato per la crittografia SSL mediante HTTPS. Agli utenti viene chiesto di considerare affidabile il certificato autofirmato quando si collegano. Puoi utilizzare un certificato affidabile e chiavi di tua proprietà. Sostituisci il file di certificato predefinito, `server.crt`, e il file della chiave `server.key` nella directory `/etc/jupyter/conf/` sul nodo master con file di certificato e di chiave personali. Utilizza le proprietà `c.JupyterHub.ssl_key` e `c.JupyterHub.ssl_cert` nel file `jupyterhub_config.py` per specificare i materiali SSL. Per ulteriori informazioni, consulta la sezione relativa alle [Security settings \(Impostazioni di sicurezza\)](#) nella documentazione di JupyterHub. Dopo aver aggiornato `jupyterhub_config.py`, riavvia il container.

Configurazione e amministrazione di JupyterHub

JupyterHub e i componenti correlati vengono eseguiti all'interno di un container Docker denominato `jupyterhub` che esegue il sistema operativo Ubuntu. Esistono diversi modi per amministrare i componenti in esecuzione all'interno del container.

Warning

Le personalizzazioni eseguite all'interno del container potrebbero non persistere in caso di riavvio del container. Ti consigliamo di creare uno script o automatizzare la configurazione del container in modo da riprodurre le personalizzazioni più facilmente.

Amministrazione mediante la riga di comando

In caso di connessione al nodo master tramite SSH, puoi inviare comandi utilizzando l'interfaccia a riga di comando (CLI) del Docker e specificando il container per nome (`jupyterhub`) o ID. Ad esempio, `sudo docker exec jupyterhub command` esegue i comandi riconosciuti dal sistema operativo o un'applicazione in esecuzione all'interno del container. Puoi utilizzare questo metodo per aggiungere utenti al sistema operativo e installare applicazioni e librerie aggiuntive all'interno del container Docker. Ad esempio, l'immagine del container predefinito include Conda per installazione di pacchetti, pertanto puoi eseguire il comando seguente sulla riga di comando del nodo master per installare un'applicazione, Keras, all'interno del container:

```
sudo docker exec jupyterhub conda install keras
```

Amministrazione inviando fasi

Le fasi sono un modo per inviare lavori a un cluster. Puoi inviare fasi all'avvio di un cluster o puoi inviare fasi a un cluster in esecuzione. I comandi eseguiti sulla riga di comando possono essere inviati come fasi mediante `command-runner.jar`. Per ulteriori informazioni, consulta la sezione relativa all'[Utilizzo di fasi mediante la CLI e la console](#) nella Guida per la gestione di Amazon EMR e [Eseguire comandi e script su un cluster Amazon EMR](#).

Ad esempio, puoi utilizzare il comando AWS CLI seguente su un computer locale per installare Keras così come è stato fatto dalla riga di comando del nodo master nell'esempio precedente:

```
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"
```

Inoltre, puoi creare uno script di una sequenza di fasi, caricare lo script in Amazon S3, quindi utilizzare `script-runner.jar` per eseguire lo script durante la creazione del cluster o l'aggiunta

dello script come una fase. Per ulteriori informazioni, consulta [Eseguire comandi e script su un cluster Amazon EMR](#). Per vedere un esempio, consulta [the section called “Esempio: Script bash per aggiungere più utenti”](#).

Amministrazione utilizzando API REST

Jupyter, JupyterHub e il proxy HTTP per JupyterHub forniscono API REST che puoi utilizzare per inviare richieste. Per inviare le richieste a JupyterHub, devi passare un token API con la richiesta. Puoi utilizzare il comando `curl` dalla riga di comando del nodo master per eseguire i comandi REST. Per ulteriori informazioni, consulta le seguenti risorse :

- [Utilizzo della REST API di JupyterHub](#) nella documentazione per JupyterHub, che include istruzioni per la generazione di token API
- [API server Jupyter Notebook](#) su GitHub
- [configurable-http-proxy](#) su GitHub

Nell'esempio seguente viene illustrato l'utilizzo della REST API per JupyterHub per ottenere un elenco di utenti. Il comando passa un token di amministrazione generato in precedenza e utilizza la porta predefinita, 9443, per JupyterHub, reindirizzando l'output a `jq` per facilitare la visualizzazione:

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \  
-H "Authorization: token $admin_token" | jq .
```

Aggiunta di utenti e amministratori Jupyter Notebook

Puoi utilizzare uno dei due metodi per gli utenti da autenticare in JupyterHub in modo che possano creare notebook e, facoltativamente, amministrare JupyterHub. Il metodo più semplice è utilizzare PAM (Pluggable Authentication Module) di JupyterHub. Inoltre, JupyterHub su Amazon EMR supporta [LDAP Authenticator Plugin per JupyterHub](#) per ottenere identità utente da un server LDAP, ad esempio un server di Microsoft Active Directory. Istruzioni ed esempi per l'aggiunta di utenti con ogni metodo di autenticazione vengono forniti in questa sezione.

JupyterHub su Amazon EMR dispone di un utente predefinito con autorizzazioni di amministratore. Il nome utente è `jovyan` e la password è `jupyter`. Ti consigliamo di sostituire l'utente con un altro che dispone di autorizzazioni amministrative. È possibile eseguire questa operazione utilizzando una fase della creazione del cluster, oppure collegandosi al nodo master quando il cluster è in esecuzione.

Argomenti

- [Utilizzo dell'autenticazione PAM](#)
- [Utilizzo dell'autenticazione LDAP](#)
- [Rappresentazione dell'utente](#)

Utilizzo dell'autenticazione PAM

La creazione di utenti PAM in JupyterHub su Amazon EMR è un processo in due fasi. La prima fase è aggiungere utenti al sistema operativo in esecuzione nel container `jupyterhub` sul nodo master e aggiungere una home directory utente corrispondente per ogni utente. La seconda fase è aggiungere questi utenti del sistema operativo come utenti JupyterHub, un processo noto come `whitelist` in JupyterHub. Dopo che sono stati aggiunti, gli utenti JupyterHub possono connettersi all'URL JupyterHub e fornire le relative credenziali del sistema operativo per l'accesso.

Quando un utente effettua l'accesso, JupyterHub apre l'istanza del server del notebook per tale utente, che viene salvata nella home directory dell'utente sul nodo master, che è `/var/lib/jupyter/home/username`. Se un'istanza del server del notebook non esiste, JupyterHub genera dinamicamente un'istanza notebook nella home directory dell'utente. Nelle seguenti sezioni viene illustrato come aggiungere utenti individualmente al sistema operativo e a JupyterHub, seguito da uno script bash rudimentale che consente di aggiungere più utenti.

Aggiunta di un utente del sistema operativo al container

Nell'esempio seguente viene utilizzato innanzitutto il comando [`useradd`](#) all'interno del container per aggiungere un singolo utente, `diego`, e creare una home directory per tale utente. Il secondo comando utilizza [`chpasswd`](#) per stabilire una password di `diego` per questo utente. I comandi vengono eseguiti sulla riga di comando del nodo master durante la connessione mediante SSH. Puoi anche eseguire questi comandi utilizzando una fase come descritto in precedenza in [Amministrazione inviando fasi](#).

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

Aggiunta di un utente JupyterHub

Puoi utilizzare il pannello Admin (Amministratore) in JupyterHub o la REST API per aggiungere utenti e amministratori, oppure solo utenti.

Aggiunta di utenti e amministratori utilizzando il pannello Admin (Amministratore) in JupyterHub

1. Connettersi al nodo master mediante SSH e accedere a `https://MasterNodeDNS:9443` con un'identità che dispone di autorizzazioni di amministratore.
2. Scegliere Control Panel (Pannello di controllo), Admin (Amministratore).
3. Scegliere User (Utente), Add Users (Aggiungi utenti) o scegliere Admin (Amministratore), Add Admins (Aggiungi amministratori).

Per aggiungere un utente utilizzando la REST API

1. Connettersi al nodo master mediante SSH e utilizzare il comando seguente sul nodo master o eseguire il comando come una fase.
2. Acquisire un token amministrativo per creare richieste API e sostituire *AdminToken* nella fase seguente con tale token.
3. Utilizzare il comando seguente, sostituendo *UserName* con un utente del sistema operativo che è stato creato all'interno del container.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

Note

Quando accedi all'interfaccia Web di JupyterHub per la prima volta, vieni aggiunto automaticamente come utente non amministratore di JupyterHub.

Esempio: Script bash per aggiungere più utenti

Lo script bash di esempio seguente lega tra loro le fasi precedenti in questa sezione per creare più utenti JupyterHub. Lo script può essere eseguito direttamente sul nodo master o può essere caricato in Amazon S3 e, successivamente, eseguito come fase.

Il primo script stabilisce una serie di nomi utente e utilizza il comando `jupyterhub token` per creare un token API per l'amministratore predefinito, `jovyan`. Quindi crea un utente del sistema operativo nel container `jupyterhub` per ogni utente, assegnando a ciascuno una password iniziale che è uguale al relativo nome utente. Infine, chiama l'operazione della REST API per creare ogni

utente in JupyterHub. Il token generato in precedenza nello script viene passato ed la risposta REST viene reindirizzata a jq per facilitare la visualizzazione.

```
# 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
```

Salva lo script in un percorso in Amazon S3, ad esempio `s3://mybucket/createjupyterusers.sh`. Quindi puoi utilizzare `script-runner.jar` per eseguirlo come una fase.

Esempio: Esecuzione dello script durante la creazione di un cluster (AWS CLI)

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.1 \
--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"]
```

Esecuzione dello script su un cluster esistente (AWS CLI)

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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"]
```

Utilizzo dell'autenticazione LDAP

LDAP è un protocollo di applicazioni per inviare query e modificare oggetti che corrispondono a risorse quali utenti e computer archiviati in un provider di servizi di directory compatibile con LDAP, ad esempio Active Directory o un server OpenLDAP. Puoi utilizzare [LDAP Authenticator Plugin per JupyterHub](#) con JupyterHub su Amazon EMR per utilizzare LDAP per l'autenticazione utente. Il plugin gestisce sessioni di accesso per utenti LDAP e fornisce informazioni utente a Jupyter. Questo consente agli utenti di connettersi a JupyterHub e notebook utilizzando le credenziali per le loro identità archiviate in un server compatibile con LDAP.

Le fasi in questa sezione illustrano le fasi seguenti per configurare e abilitare LDAP utilizzando LDAP Authenticator Plugin per JupyterHub. Eseguire le fasi mentre si è connessi alla riga di comando del nodo master. Per ulteriori informazioni, consulta [Connessione al nodo master e ai server notebook](#).

1. Crea un file di configurazione LDAP con informazioni sul server LDAP, ad esempio l'indirizzo IP host, la porta, i nomi di binding e così via.
2. Modifica `/etc/jupyter/conf/jupyterhub_config.py` per abilitare LDAP Authenticator Plugin per JupyterHub.
3. Crea ed esegui uno script che configura LDAP all'interno del container `jupyterhub`.
4. Esegui query su LDAP per utenti e crea home directory all'interno del container per ogni utente. JupyterHub richiede home directory per ospitare notebook.
5. Esegui uno script che riavvia JupyterHub.

⚠ Important

Prima di configurare LDAP, verifica l'infrastruttura di rete per garantire che il server LDAP e il nodo master del cluster siano in grado di comunicare come richiesto. TLS utilizza in genere la porta 389 su una connessione TCP semplice. Se la connessione LDAP utilizza SSL, la porta TCP nota per SSL è 636.

Creazione del file di configurazione LDAP

Nell'esempio seguente vengono utilizzati i seguenti valori di configurazione del segnaposto. Sostituiscili con i parametri che soddisfano l'implementazione.

- Il server LDAP esegue la versione 3 ed è disponibile sulla porta 389. Questa è la porta non SSL standard per LDAP.
- Il nome distinto di base (DN) è `dc=example, dc=org`.

Utilizza un editor di testo per creare il file [ldap.conf](#), con contenuti simili ai seguenti. Utilizza valori appropriate per l'implementazione LDAP. Sostituisci *host* con l'indirizzo IP o il nome host risolvibile del server LDAP.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

Abilitazione di LDAP Authenticator Plugin per JupyterHub

Utilizza un editor di testo per modificare il file `/etc/jupyter/conf/jupyterhub_config.py` e aggiungere proprietà [ldapauthenticator](#) simili alle seguenti. Sostituisci *host* con l'indirizzo IP o il nome host risolvibile del server LDAP. L'esempio presuppone che gli oggetti utente si trovino all'interno di un'unità organizzativa (ou) denominata *people* e utilizza i componenti nome distinto definiti in precedenza utilizzando `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'
```

Configurazione di LDAP all'interno del container

Utilizza un editor di testo per creare uno script bash con il seguente contenuto:

```
#!/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"
```

Salva lo script nel nodo master, quindi esegilo dalla riga di comando del nodo master. Ad esempio, con lo script salvato come `configure_ldap_client.sh`, rendi il file eseguibile:

```
chmod +x configure_ldap_client.sh
```


Ed esegui lo script:

```
./configure_ldap_client.sh
```

Aggiunta di attributi ad Active Directory

Per trovare ogni utente e creare la voce appropriata nel database, il container docker JupyterHub richiede le seguenti proprietà UNIX per il corrispondente oggetto utente in Active Directory. Per ulteriori informazioni, consulta la sezione relativa a [Come si può continuare a modificare gli attributi LDAP RFC 2307 ora che il plug-in degli attributi Unix non è più disponibile per lo snap-in MMC di utenti e computer di Active Directory nell'articolo sul Chiarimento sullo stato di Identity Management per Unix \(IDMU\) e sul ruolo del server NIS in Windows Server 2016 Technical Preview e oltre.](#)

- `homeDirectory`

Questo è il percorso alla home directory dell'utente, in genere `/home/username`.

- `gidNumber`

Si tratta di un valore maggiore di 60000 che non è già utilizzato da un altro utente. Seleziona il file `etc/passwd` per i gid in uso.

- `uidNumber`

Si tratta di un valore maggiore di 60000 che non è già utilizzato da un altro gruppo. Seleziona il file `etc/group` per gli uid in uso.

- `uid`

È uguale al *nome utente*.

Creazione di home directory dell'utente

JupyterHub richiede home directory all'interno del container per autenticare gli utenti LDAP e archiviare i dati dell'istanza. Nell'esempio seguente vengono illustrati due utenti, shirley e diego, nella directory LDAP.

La prima fase è eseguire una query sul server LDAP per ogni informazione id utente e id gruppo dell'utente utilizzando [ldapsearch](#) come mostrato nell'esempio seguente, sostituendo *host* con l'indirizzo IP o il nome host risolvibile del server 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
```

Il comando `ldapsearch` restituisce una risposta in formato LDIF il cui aspetto è simile al seguente per gli utenti `shirley` e `diego`.

```
# extended LDIF  
  
# LDAPv3  
# base <ou=people,dc=example,dc=org> with scope subtree  
# filter: (objectclass=*)  
# requesting: uidNumber gidNumber sn  
  
# people, example.org  
dn: ou=people,dc=example,dc=org  
  
# diego, people, example.org  
dn: cn=diego,ou=people,dc=example,dc=org  
sn: B  
uidNumber: 1001  
gidNumber: 100  
  
# shirley, people, example.org  
dn: cn=shirley,ou=people,dc=example,dc=org  
sn: A  
uidNumber: 1002  
gidNumber: 100  
  
# search result  
search: 2  
result: 0 Success  
  
# numResponses: 4  
# numEntries: 3
```

Utilizzando le informazioni della risposta, esegui comandi all'interno del container per creare una home directory per ogni nome comune dell'utente (`cn`). Utilizza `uidNumber` e `gidNumber` per correggere la proprietà per la home directory di tale utente. I comandi dell'esempio seguente consentono di eseguire questa operazione per l'utente *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

L'autenticatore LDAP per JupyterHub non supporta la creazione di utenti locali. Per ulteriori informazioni, consulta [Nota sulla configurazione di LDAP Authenticator per la creazione di utenti locali](#).

Per creare manualmente un utente locale, utilizzare il comando seguente.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

Riavvio del container JupyterHub

Esegui i comandi seguenti per riavviare il container jupyterhub:

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

Rappresentazione dell'utente

Un processo Spark in esecuzione all'interno di un notebook Jupyter passa attraverso più applicazioni durante la propria esecuzione su Amazon EMR. Ad esempio, il codice PySpark3 che un utente esegue all'interno di Jupyter viene ricevuto da Sparkmagic, che utilizza una richiesta HTTP POST per inviarlo a Livy, che quindi crea un processo Spark da eseguire sul cluster utilizzando YARN.

Per impostazione predefinita, le operazioni YARN inviate in questo modo vengono eseguite come utente `livy`, indipendentemente dall'utente che ha avviato il processo. Impostando la rappresentazione degli utenti si può fare in modo che l'ID utente dell'utente del notebook sia anche l'utente associato al processo YARN. Anziché avere processi avviati da entrambi `shirley` e `diego` associati con l'utente `livy`, i processi avviati da ciascun utente sono associati a `shirley` e `diego` rispettivamente. In questo modo è possibile effettuare un audit dell'utilizzo di Jupyter e gestire le applicazioni all'interno della propria organizzazione.

Questa configurazione è supportata solo quando le chiamate da Sparkmagic a Livy non sono autenticate. Le applicazioni che forniscono un livello di autenticazione o di proxy tra le applicazioni

Hadoop e Livy (ad esempio Apache Knox Gateway) non sono supportate. La procedura per configurare la rappresentazione degli utenti in questa sezione presuppone che JupyterHub and Livy siano in esecuzione sullo stesso nodo master. Se la tua applicazione ha cluster separati, [Fase 3: crea le directory home HDFS per gli utenti](#) deve essere modificato in modo che le directory HDFS vengano create nel nodo master Livy.

Fasi per configurare la rappresentazione degli utenti

- [Fase 1: configurazione di Livy](#)
- [Fase 2: aggiungere gli utenti](#)
- [Fase 3: crea le directory home HDFS per gli utenti](#)

Fase 1: configurazione di Livy

È possibile utilizzare le classificazioni di configurazioni `livy-conf` e `core-site` al momento della creazione di un cluster per consentire la rappresentazione dell'utente Livy come mostrato nel seguente esempio. Salva la classificazione di configurazione come JSON, quindi fai riferimento a essa al momento della creazione del cluster, oppure specifica la configurazione della classificazione inline. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]
```

Fase 2: aggiungere gli utenti

Aggiungi gli utenti JupyterHub utilizzando PAM o LDAP. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione PAM](#) e [Utilizzo dell'autenticazione LDAP](#).

Fase 3: crea le directory home HDFS per gli utenti

Ti sei connesso al nodo master per creare gli utenti. Mentre sei ancora connesso al nodo master, copia i contenuti di seguito e salvali in un file di script. Lo script crea directory home HDFS per ogni utente JupyterHub sul nodo master. Lo script presuppone che stai utilizzando l'ID utente dell'amministratore di default, *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
```

Installazione di kernel e librerie aggiuntive

Quando crei un cluster con JupyterHub su Amazon EMR, il kernel Python 3 predefinito per Jupyter e i kernel PySpark e Spark per Sparkmagic vengono installati nel container Docker. Puoi installare kernel aggiuntivi. Puoi anche installare librerie e pacchetti aggiuntivi e quindi importarli per la shell appropriata.

Installazione di un kernel

Kernel vengono installati all'interno del container Docker. Il modo più semplice per effettuare questa operazione è creare uno script bash con i comandi di installazione, salvarlo nel nodo master, quindi utilizzare il comando `sudo docker exec jupyterhub script_name` per eseguire lo script all'interno del container `jupyterhub`. Nell'esempio seguente lo script installa il kernel, quindi installa alcune librerie per tale kernel nel nodo master per consentire di importarle in seguito utilizzando il kernel in Jupyter.

```
#!/bin/bash

# Install Python 2 kernel
conda create -n py27 python=2.7 anaconda
source /opt/conda/envs/py27/bin/activate
apt-get update
apt-get install -y gcc
/opt/conda/envs/py27/bin/python -m pip install --upgrade ipykernel
/opt/conda/envs/py27/bin/python -m ipykernel install

# Install libraries for Python 2
/opt/conda/envs/py27/bin/pip install paramiko nltk scipy numpy scikit-learn pandas
```

Per installare il kernel e le librerie all'interno del container, apri una connessione del terminale al nodo master, salva lo script in `/etc/jupyter/install_kernels.sh` ed esegui il comando seguente nella riga di comando del nodo master:

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

Utilizzo di librerie e installazione di librerie aggiuntive

Un set di base di librerie di machine learning e data science per Python 3 è preinstallato con JupyterHub su Amazon EMR. Puoi utilizzare `sudo docker exec jupyterhub bash -c "conda list"` e `sudo docker exec jupyterhub bash -c "pip freeze"`.

Se un processo Spark richiede librerie sui nodi di lavoro, ti consigliamo di utilizzare un'operazione di bootstrap per eseguire uno script per installare le librerie durante la creazione del cluster. Le operazioni di bootstrap vengono eseguite su tutti i nodi del cluster durante il processo di creazione del cluster; ciò semplifica l'installazione. Se installi librerie su nodi principali/di lavoro dopo che un cluster è in esecuzione, l'operazione è più complessa. In questa sezione viene fornito un esempio di programma Python che mostra come installare queste librerie.

L'operazione di bootstrap e il programma Python di esempio mostrati in questa sezione utilizzano uno script bash salvato in Amazon S3 per installare le librerie su tutti i nodi.

Lo script cui si fa riferimento negli esempi seguenti utilizza pip per installare paramiko, nltk, scipy, scikit-learn e pandas per il kernel Python 3:

```
#!/bin/bash

sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

Dopo aver creato lo script, caricalo in un percorso in Amazon S3, ad esempio `s3://mybucket/install-my-jupyter-libraries.sh`. Per ulteriori informazioni, consulta [Caricamento di oggetti](#) nella Guida per l'utente di Amazon Simple Storage Service per utilizzare lo script nell'operazione di bootstrap o nel programma Python.

Per specificare un'operazione di bootstrap che installa librerie su tutti i nodi durante la creazione di un cluster utilizzando AWS CLI

1. Crea uno script simile all'esempio precedente e salvalo in un percorso in Amazon S3. Viene utilizzato l'esempio `s3://mybucket/install-my-jupyter-libraries.sh`.
2. Crea il cluster con JupyterHub e utilizza l'argomento Path dell'opzione `--bootstrap-actions` per specificare il percorso dello script come nell'esempio seguente:

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.1 \
--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
```

Per specificare un'operazione di bootstrap che installa librerie su tutti i nodi durante la creazione di un cluster utilizzando la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Specificare le impostazioni per Software and Steps (Software e fasi) e Hardware come appropriato per l'applicazione.
4. Nella General Cluster Settings (Impostazioni generali del cluster), espandere Bootstrap Actions (Operazioni di bootstrap).
5. Per Add bootstrap action (Aggiungi operazione di bootstrap), selezionare Custom action (Operazione personalizzata), Configure and add (Configura e aggiungi).
6. Per Name (Nome), inserire un nome descrittivo. Per Script location (Percorso script), inserisci il percorso in Amazon S3 dello script (nell'esempio usato è *s3://mybucket/install-my-jupyter-libraries.sh*). Lasciare Optional arguments (Argomenti opzionali) vuoto e scegliere Add (Aggiungi).
7. Specificare altre impostazioni per il cluster e scegliere Next (Successivo).
8. Specificare impostazioni di sicurezza e scegliere Create cluster (Crea cluster).

Example Installazione di librerie su nodi principali di un cluster in esecuzione

Dopo aver installato le librerie sul nodo master dall'interno di Jupyter, puoi installare librerie su nodi principali in esecuzione in diversi modi. Nell'esempio seguente viene mostrato un programma Python scritto per l'esecuzione su un computer locale. Quando esegui il programma Python localmente, `AWS-RunShellScript` di AWS Systems Manager viene utilizzato per eseguire lo script di esempio, illustrato in precedenza in questa sezione, che consente di installare librerie sui nodi principali del cluster.

```
import argparse  
import time
```



```
import boto3

def install_libraries_on_core_nodes(cluster_id, script_path, emr_client, ssm_client):
    """
    Copies and runs a shell script on the core nodes in the cluster.

    :param cluster_id: The ID of the cluster.
    :param script_path: The path to the script, typically an Amazon S3 object URL.
    :param emr_client: The Boto3 Amazon EMR client.
    :param ssm_client: The Boto3 AWS Systems Manager client.
    """
    core_nodes = emr_client.list_instances(
        ClusterId=cluster_id, InstanceGroupTypes=["CORE"]
    )["Instances"]
    core_instance_ids = [node["Ec2InstanceId"] for node in core_nodes]
    print(f"Found core instances: {core_instance_ids}.")

    commands = [
        # Copy the shell script from Amazon S3 to each node instance.
        f"aws s3 cp {script_path} /home/hadoop",
        # Run the shell script to install libraries on each node instance.
        "bash /home/hadoop/install_libraries.sh",
    ]
    for command in commands:
        print(f"Sending '{command}' to core instances...")
        command_id = ssm_client.send_command(
            InstanceIds=core_instance_ids,
            DocumentName="AWS-RunShellScript",
            Parameters={"commands": [command]},
            TimeoutSeconds=3600,
        )["Command"]["CommandId"]
        while True:
            # Verify the previous step succeeded before running the next step.
            cmd_result = ssm_client.list_commands(CommandId=command_id)["Commands"][0]
            if cmd_result["StatusDetails"] == "Success":
                print(f"Command succeeded.")
                break
            elif cmd_result["StatusDetails"] in ["Pending", "InProgress"]:
                print(f"Command status is {cmd_result['StatusDetails']}, waiting...")
                time.sleep(10)
            else:
                print(f"Command status is {cmd_result['StatusDetails']}, quitting.")
                raise RuntimeError(
```

```

        f"Command {command} failed to run. "
        f"Details: {cmd_result['StatusDetails']}"
    )

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("cluster_id", help="The ID of the cluster.")
    parser.add_argument("script_path", help="The path to the script in Amazon S3.")
    args = parser.parse_args()

    emr_client = boto3.client("emr")
    ssm_client = boto3.client("ssm")

    install_libraries_on_core_nodes(
        args.cluster_id, args.script_path, emr_client, ssm_client
    )

if __name__ == "__main__":
    main()

```

Cronologia dei rilasci di JupyterHub

Nelle tabelle seguenti sono elencati la versione di JupyterHub inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni relative alla versione di JupyterHub

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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-

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con JupyterHub
		yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di JupyterHub	Componenti installati con 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 consente l'interazione mediante un'interfaccia REST con un cluster EMR in cui Spark è in esecuzione. Puoi utilizzare l'interfaccia REST o una libreria di client RPC per inviare processi Spark o frammenti di codice Spark, recuperare i risultati in modo sincrono o asincrono e gestire Spark Context. Per ulteriori informazioni, consulta il [sito Web di Apache Livy](#). Livy è incluso in Amazon EMR 5.9.0 e versioni successive.

Per accedere all'interfaccia Web Livy, impostare un tunnel SSH al nodo master e una connessione proxy. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce Web ospitate su cluster Amazon EMR](#).

La seguente tabella indica la versione di Livy inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Livy.

Per la versione dei componenti installati con Livy in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Livy per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.14.0	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

La seguente tabella indica la versione di Livy inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Livy.

Per la versione dei componenti installati con Livy in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Livy per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.36.1	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

Argomenti

- [Abilitazione di HTTPS con Apache Livy](#)
- [Cronologia dei rilasci di Livy](#)

Abilitazione di HTTPS con Apache Livy

1. Esegui il provisioning di un cluster Amazon EMR con crittografia di transito abilitata. Per ulteriori informazioni sulla crittografia, consulta [Crittografia dei dati a riposo e in transito](#).
2. Crea un file denominato `livy_ssh.sh`, con i seguenti contenuti.

```
#!/bin/bash
```

```

KEYSTORE_FILE=`awk '/ssl.server.keystore.location/{getline; print}' /etc/hadoop/
conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEYSTORE_PASS=`awk '/ssl.server.keystore.password/{getline; print}' /etc/hadoop/
conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEY_PASS=`awk '/ssl.server.keystore.keypassword/{getline; print}' /etc/hadoop/conf/
ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`

echo "livy.keystore $KEYSTORE_FILE
livy.keystore.password $KEYSTORE_PASS
livy.key-password $KEY_PASS" | sudo tee -a /etc/livy/conf/livy.conf >/dev/null

sudo systemctl restart livy-server.service

```

3. Esegui lo script seguente come fase di Amazon EMR. Questo script modifica `/etc/livy/conf/livy.conf` per attivare SSL.

```

--steps '[{"Args":["s3://DOC-EXAMPLE-BUCKET/
livy_ssl.sh"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"s3://
us-east-1.elasticmapreduce/libs/script-runner/script-
runner.jar","Properties":"","Name":"Custom JAR"}]'

```

4. Riavvia il servizio Apache Livy in modo che le modifiche diventino effettive. Per riavviare Apache Livy, consulta [Arresto e riavvio di processi](#).
5. Verifica che i client possano comunicare utilizzando l'HTTPS. Per inviare un processo, ad esempio, esegui il codice seguente.

```

curl -k -X POST --data '{"file": "local:///usr/lib/spark/examples/jars/spark-
examples.jar",
"className": "org.apache.spark.examples.SparkPi"}' \
-H "Content-Type: application/json" \
https://EMR_Master_Node_Host:8998/batches

```

Se HTTPS è stato abilitato correttamente, Livy invia un messaggio di risposta indicante che il comando è stato accettato e che il processo batch è stato inviato.

```

{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo"
{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ","\nstderr: ","\nYARN
Diagnostics: "]}

```

Cronologia dei rilasci di Livy

La seguente tabella indica la versione di Livy inclusa in ogni rilascio di Amazon EMR, insieme ai componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Livy

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.14.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.13.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server,

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
		spark-on-yarn, spark-yarn-slave, livy-server, nginx
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

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.11.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.10.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.10.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.9.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.9.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.8.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.8.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.7.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.36.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.6.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.35.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.5.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.4.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.3.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.3.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.2.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.2.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.1.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.1.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-6.0.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.0.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.34.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.33.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.33.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.32.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.32.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.31.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.31.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.2	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.30.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.29.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.28.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.28.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.27.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.27.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.26.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.25.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.24.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.24.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.23.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.23.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.22.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.21.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.21.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.21.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.20.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.20.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.19.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.19.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.18.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.18.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.17.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.17.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.17.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.16.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.16.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.15.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.15.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.14.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.14.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.14.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.13.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.13.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.12.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.12.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.4	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.11.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.11.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.10.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.10.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Etichetta di rilascio di Amazon EMR	Versione di Livy	Componenti installati con Livy
emr-5.9.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.9.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Apache MXNet

Apache MXNet è una libreria di accelerazione concepita per la creazione di reti neurali e altre applicazioni di Deep Learning. MXNet automatizza i flussi di lavoro comuni e ottimizza i calcoli numerici. MXNet ti consente di progettare architetture di rete neurale senza preoccuparti di implementare calcoli di basso livello come le operazioni di algebra lineare. MXNet è incluso in Amazon EMR 5.10.0 e versioni successive.

Per ulteriori informazioni, consulta il [sito Web di Apache MXNet](#).

La seguente tabella indica la versione di MXNet inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con MXNet.

Per la versione dei componenti installati con MXNet in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di MXNet per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.14.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

La seguente tabella indica la versione di MXNet inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con MXNet.

Per la versione dei componenti installati con MXNet in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di MXNet per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.36.1	MXNet 1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Cronologia dei rilasci di MXNet

La seguente tabella indica la versione di MXNet inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di MXNet

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.14.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.13.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.12.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.11.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-6.10.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.9.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.8.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.1	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.36.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.6.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.35.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.5.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.4.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.3.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.1.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-6.0.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-6.0.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.34.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.33.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-5.32.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.31.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.30.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-5.29.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.27.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.25.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.24.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.1	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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.23.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-5.21.2	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.21.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.20.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
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

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.18.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.2	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.17.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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
emr-5.16.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.16.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.14.2	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.13.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.13.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.3	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.12.2	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.11.4	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.3	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.2	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con 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
emr-5.10.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet

Etichetta di rilascio di Amazon EMR	Versione di MXNet	Componenti installati con MXNet
emr-5.10.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Apache Oozie

Utilizza il pianificatore di workflow Apache Oozie per gestire e coordinare i processi Hadoop. Per ulteriori informazioni, vedi <http://oozie.apache.org/>.

L'interfaccia Web nativa di Oozie non è supportata su Amazon EMR. Per utilizzare un'interfaccia front-end per Oozie, prova l'applicazione Hue Oozie. Per ulteriori informazioni, consulta [Hue](#). Oozie è incluso con il rilascio di Amazon EMR versione 5.0.0 e successive. Nelle versioni precedenti, Oozie è incluso come applicazione sandbox. Per ulteriori informazioni, consulta [Versioni del rilascio di Amazon EMR 4.x](#).

Se si utilizza un'AMI Amazon Linux personalizzata basata su un'AMI Amazon Linux con una data di creazione 2018-08-11, il server Oozie non si avvia. Se si utilizza Oozie, è possibile creare un'AMI personalizzata basata su un ID AMI Amazon Linux con una data di creazione diversa. Puoi utilizzare il comando AWS CLI seguente per restituire un elenco di ID immagine per tutte le AMI Amazon Linux HVM con una versione 2018.03, insieme alla data di rilascio, in modo da poter scegliere un'AMI Linux appropriata come base. Sostituisci MyRegion con l'identificatore della Regione, ad esempio 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
```

La seguente tabella indica la versione di Oozie inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Oozie.

Per la versione dei componenti installati con Oozie in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Oozie per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.14.0	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-

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
		hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, hadoop-yarn- timeline-server, oozie-client, oozie-server, tez-on-yarn, tez- on-worker

La seguente tabella indica la versione di Oozie inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Oozie.

Per la versione dei componenti installati con Oozie in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Oozie per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.36.1	Oozie 5.2.1	emrfs, emr-ddb, emr-goodi- es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma- pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, hadoop-yarn- timeline-server, oozie-client, oozie-server, tez-on-yarn

Argomenti

- [Utilizzo di Oozie con un database remoto in Amazon RDS](#)
- [Configurazione della versione Java per Oozie](#)
- [Cronologia dei rilasci di Oozie](#)

Utilizzo di Oozie con un database remoto in Amazon RDS

Per impostazione predefinita, le informazioni utente Oozie e le cronologie delle query vengono archiviate in un database MySQL locale sul nodo master. In alternativa, puoi creare uno o più cluster abilitati Oozie utilizzando una configurazione archiviata in Amazon S3 e un database MySQL in Amazon Relational Database Service (Amazon RDS). In questo modo puoi preservare le informazioni utente e la cronologia delle query create da Oozie senza mantenere il cluster Amazon EMR in esecuzione. Ti consigliamo di usare la crittografia lato server Amazon S3 per archiviare il file di configurazione.

Innanzitutto, crea il database remoto per Oozie.

Creazione del database MySQL esterno

1. Apri la console di Amazon RDS all'indirizzo <https://console.aws.amazon.com/rds/>.
2. Seleziona Launch a DB instance (Avvia istanza database).
3. Scegliere MySQL, quindi selezionare Seleziona.
4. Lasciare la selezione predefinita di Multi-AZ Deployment and Provisioned IOPS Storage (Implementazione Multi-AZ e storage IOPS assegnato) e fare clic su Next (Successivo).
5. Lasciare le impostazioni predefinite per Instance Specifications (Specifiche delle istanze), specificare Settings (Impostazioni) e fare clic su Next (Successivo).
6. Nella pagina Configure Advanced Settings (Configura impostazioni avanzate), scegli un gruppo di sicurezza e nomi di database adeguati. Il gruppo di sicurezza utilizzato deve consentire almeno l'accesso TCP in entrata per la porta 3306 dal nodo master del cluster. Se a questo punto il cluster non è ancora stato creato, puoi consentire a tutti gli host di connettersi alla porta 3306 e modificare il gruppo di sicurezza dopo che il cluster è stato avviato. Seleziona Avvia istanza database.
7. Dal pannello di controllo RDS, scegliere Instances (Istanze) e selezionare l'istanza appena creata. Quando il database è disponibile, prendere nota di namedb, nome utente, password e nome host dell'istanza RDS. Utilizzare questa informazione quando si crea e configura il cluster.

Per specificare un database MySQL esterno per Oozie quando si avvia un cluster utilizzando AWS CLI

Per specificare un database MySQL esterno per Oozie quando si avvia un cluster utilizzando AWS CLI, utilizzare le informazioni annotate durante la creazione dell'istanza RDS per configurare `oozie-site` con un oggetto di configurazione

Note

Puoi creare più cluster che utilizzano lo stesso database esterno, ma ogni cluster condividerà la cronologia delle query e le informazioni utente.

- Utilizzando AWS CLI, creare un cluster con Oozie installato, utilizzando il database esterno creato e facendo riferimento a un file di configurazione con una classificazione della configurazione per Oozie che specifica le proprietà del database. L'esempio seguente consente di creare un cluster con Oozie installato, facendo riferimento a un file di configurazione in Amazon S3, `myConfig.json`, che specifica la configurazione del database.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --release-label emr-5.36.1 --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
```

Contenuti di esempio del file `myConfig.json` sono mostrati di seguito. Sostituire *l'URL JDBC*, *il nome utente* e la *password* con l'URL JDBC, il nome utente e la password dell'istanza RDS.

⚠ Important

L'URL JDBC deve includere il nome del database come suffisso. Ad esempio, jdbc:mysql://oozie-external-db.xxxxxxxxxx.us-east-1.rds.amazonaws.com:3306/dbname.

```
[{
  "Classification": "oozie-site",
  "Properties": {
    "oozie.service.JPAService.jdbc.driver": "org.mariadb.jdbc.Driver",
    "oozie.service.JPAService.jdbc.url": "JDBC URL",
    "oozie.service.JPAService.jdbc.username": "username",
    "oozie.service.JPAService.jdbc.password": "password"
  },
  "Configurations": []
}]
```

Configurazione della versione Java per Oozie

Oozie esegue più processi della macchina virtuale Java (JVM). Questa pagina spiega come configurare la versione di Java per ogni processo.

- **Server Oozie:** imposta JAVA_HOME nella classificazione di oozie-env per aggiornare la versione Java per EmbeddedOozieServer.
- **Oozie Launcher AM:** Oozie Launcher AM è un processo MR a mappatura singola che richiama le librerie client delle applicazioni appropriate come Hadoop e Hive. Se non diversamente configurato, le versioni di runtime per Oozie Launcher AM sono le stesse dei runtime Java per Hadoop nel cluster EMR. Per configurare il runtime Java per Oozie Launcher AM, imposta la seguente proprietà nel workflow.xml per il processo:

```
<property>
  <name>mapred.child.env</name>
  <value>JAVA_HOME=/path/to/JAVA_HOME</value>
</property>
```

Questa proprietà assicura che Oozie Launcher AM per il processo Oozie venga eseguito sulla versione Java specificata, anziché sulla versione Java impostata in Hadoop.

- Eseguitibile del client dell'applicazione: poiché Oozie Launcher AM richiama il client dell'applicazione per impostazione predefinita, il runtime Java per l'eseguitibile del client è lo stesso di Oozie Launcher AM.
- Applicazioni avviate da un processo Oozie: se non diversamente specificato, le versioni di runtime per le JVM effettive dell'applicazione che vengono avviate da un processo Oozie sono le stesse dei runtime Java per Hadoop nel cluster EMR. A seconda del tipo di operazione del flusso di lavoro di Oozie utilizzata per avviare l'applicazione in un processo Oozie (operazione Spark o Hive), puoi aggiornare il runtime Java predefinito per le JVM dell'applicazione effettiva nel `workflow.xml` per il processo Oozie.

Cronologia dei rilasci di Oozie

La seguente tabella indica la versione di Oozie inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Oozie

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
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,

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
		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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.11.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.11.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.9.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.9.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.8.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.8.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.7.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.36.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.35.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.5.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.4.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.3.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.3.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.2.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.2.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.1.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.1.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.0.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-6.0.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.34.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.33.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.33.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.32.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.32.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.31.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.31.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.30.2	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.30.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.30.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.29.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.28.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.28.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.27.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.27.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.26.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.25.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.24.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.24.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.23.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.23.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.22.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.21.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.20.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.20.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.19.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.19.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.18.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.18.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.17.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.17.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.17.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.16.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.16.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.15.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.15.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.14.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.13.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.13.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.12.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.12.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.11.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.11.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.11.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.10.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.10.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.9.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.9.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.8.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.8.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.7.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.7.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.6.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.6.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.5.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.5.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.4.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.4.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.3.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.3.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.3.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.2.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.1.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.1.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Oozie	Componenti installati con Oozie
emr-5.0.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Apache Phoenix

Apache Phoenix viene utilizzato per OLTP e analisi operative, consentendo di utilizzare query SQL standard e API JDBC per lavorare con un archivio di backup Apache HBase. Per ulteriori informazioni, consulta [Phoenix in 15 minuti o meno](#). Phoenix è incluso nel rilascio di Amazon EMR versione 4.7.0 e successive.

Se effettui l'aggiornamento da una versione precedente di Amazon EMR ad Amazon EMR versione 5.4.0 o successive e utilizzi un'indicizzazione secondaria, effettua l'aggiornamento degli indici locali come descritto nella [documentazione su Apache Phoenix](#). Amazon EMR rimuove le configurazioni richieste dalla classificazione `hbase-site`, ma gli indici devono essere ripopolati. Gli aggiornamenti online e offline degli indici sono supportati. Gli aggiornamenti online sono quelli predefiniti, il che significa che gli indici vengono ripopolati durante la fase di inizializzazione dai client Phoenix della versione 4.8.0 o più grande. Per specificare gli aggiornamenti offline, imposta la configurazione `phoenix.client.localIndexUpgrade` su falso nella classificazione `phoenix-site` e poi SSH sul nodo master per eseguire `psql [zookeeper] -1`.

La seguente tabella indica la versione di Phoenix inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Phoenix.

Per la versione dei componenti installati con Phoenix in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Phoenix per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.14.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
		timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

La seguente tabella indica la versione di Phoenix inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Phoenix.

Per la versione dei componenti installati con Phoenix in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Phoenix per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.36.1	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-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server,

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
		zookeeper-client, zookeeper-server

Argomenti

- [Creazione di un cluster con Phoenix](#)
- [Client Phoenix](#)
- [Cronologia dei rilasci di Phoenix](#)

Creazione di un cluster con Phoenix

Phoenix può essere installato scegliendo l'applicazione al momento della creazione di un cluster nella console o utilizzando la AWS CLI. Le seguenti procedure ed esempi mostrano come creare un cluster con Phoenix e HBase. Per ulteriori informazioni sulla creazione di cluster tramite la console, tra cui le Advanced Options (Opzioni avanzate), consulta la sezione relativa alla [Pianificazione e configurazione dei cluster](#) della Guida per la gestione di Amazon EMR.

Per avviare un cluster con Phoenix installato utilizzando Quick Options (Opzioni rapide) per la creazione di un cluster nella console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegliere Create cluster (Crea cluster) per utilizzare l'opzione Quick Create (Creazione rapida).
3. Per Software Configuration (Configurazione software) scegliere la release più recente appropriata per l'applicazione. Phoenix appare come opzione solo quando è selezionato Amazon Release Version emr-4.7.0 o una versione successiva.
4. Per Applications (Applicazioni) scegliere la seconda opzione, HBase: HBase **ver** con Ganglia **ver**, Hadoop **ver**, Hive **ver**, Hue **ver**, Phoenix **ver** e ZooKeeper **ver**.
5. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Nell'esempio seguente viene avviato un cluster con Phoenix installato utilizzando le impostazioni di configurazione predefinite.

Per avviare un cluster con Phoenix e HBase usando la AWS CLI

- Creare il cluster con il seguente comando:

```
aws emr create-cluster --name "Cluster with Phoenix" --release-label emr-5.36.1 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Personalizzazione delle configurazioni Phoenix

Durante la creazione di un cluster, è possibile configurare Phoenix impostando i valori di `hbase-site.xml` usando la classificazione di configurazione `hbase-site`.

Per ulteriori informazioni, consulta [Configurazione e regolazione](#) nella documentazione di Phoenix.

L'esempio seguente mostra l'utilizzo di un file JSON memorizzato in Amazon S3 per specificare il valore `false` per la proprietà `phoenix.schema.dropMetaData`. È possibile specificare più proprietà per una singola classificazione. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#). Il comando `create-cluster` fa quindi riferimento al file JSON come parametro `--configurations`.

Il contenuto del file JSON salvato in `/mybucket/myfolder/myconfig.json` è il seguente.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "phoenix.schema.dropMetaData": "false"  
    }  
  }  
]
```

]

Il comando `create cluster` che fa riferimento al file JSON viene visualizzato nell'esempio seguente.

```
aws emr create-cluster --release-label emr-5.36.1 --applications Name=Phoenix \  
Name=HBase --instance-type m5.xlarge --instance-count 2 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myconfig.json
```

Note

La richiesta di riconfigurazione per qualsiasi classificazione di configurazione Phoenix è supportata solo in Amazon EMR versione 5.23.0 e successive e non in Amazon EMR versione 5.21.0 o 5.22.0. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#)

Client Phoenix

Ci si connette a Phoenix utilizzando un client JDBC costruito con dipendenze complete oppure il "thin client" che utilizza Phoenix Query Server e che può essere eseguito solo su un nodo master di un cluster (ad esempio, utilizzando un client SQL, un passo, una riga di comando, l'inoltro della porta SSH, ecc. Quando si usa il client JDBC "fat", deve comunque avere accesso a tutti i nodi del cluster perché si connette direttamente ai servizi HBase. Il client Phoenix "thin" deve solo accedere al server query Phoenix a una porta predefinita 8765. Vi sono diversi [script](#) all'interno di Phoenix che utilizzano tali client.

Utilizzo di una fase Amazon EMR per eseguire query utilizzando Phoenix

La seguente procedura ripristina una snapshot da HBase e utilizza tali dati per eseguire una query Phoenix. È possibile estendere questo esempio o creare un nuovo script che sfrutti i client di Phoenix per soddisfare le proprie esigenze.

1. Creare un cluster con Phoenix installato, utilizzando il seguente comando:

```
aws emr create-cluster --name "Cluster with Phoenix" --log-uri s3://myBucket/  
myLogFolder --release-label emr-5.36.1 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  

```

```
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

2. Crea e in seguito carica i seguenti file in Amazon S3:

copySnapshot.sh

```
sudo su hbase -s /bin/sh -c 'hbase snapshot export \  
-D hbase.rootdir=s3://us-east-1.elasticmapreduce.samples/hbase-demo-customer-data/  
snapshot/ \  
-snapshot customer_snapshot1 \  
-copy-to hdfs://masterDNSName:8020/user/hbase \  
-mappers 2 -chuser hbase -chmod 700'
```

runQuery.sh

```
aws s3 cp s3://myBucket/phoenixQuery.sql /home/hadoop/  
/usr/lib/phoenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/  
phoenixQuery.sql
```

phoenixQuery.sql

Note

Quando si utilizza Amazon EMR versione 5.26.0 e successive, è sufficiente includere COLUMN_ENCODED_BYTES=0 nell'esempio seguente.

```
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";
```

Utilizzare la AWS CLI per inviare i file al bucket S3:

```
aws s3 cp copySnapshot.sh s3://myBucket/
aws s3 cp runQuery.sh s3://myBucket/
aws s3 cp phoenixQuery.sql s3://myBucket/
```

3. Creazione di una tabella utilizzando i seguenti passaggi inviati al cluster creato nella Fase 1:

createTable.json

```
[
  {
    "Name": "Create HBase Table",
    "Args": ["bash", "-c", "echo '$'create \"customer\", \"address\", \"cc\", \"contact
\" | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./createTable.json
```

4. Usare `script-runner.jar` per eseguire lo script `copySnapshot.sh` precedentemente caricato nel bucket S3:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="HBase Copy Snapshot",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/copySnapshot.sh"]
```

Viene eseguito un processo di MapReduce per copiare i dati dello snapshot negli HDFS del cluster.

5. Ripristinare lo snapshot copiato nel cluster utilizzando i seguenti passaggi:

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. Usare `script-runner.jar` per eseguire lo script `runQuery.sh` precedentemente caricato nel bucket S3:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Phoenix Run Query",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/runQuery.sh"]
```

La query viene eseguita e restituisce i risultati al `stdout` della fase. Il completamento della fase può richiedere alcuni minuti.

7. Ispeziona i risultati del passo `stdout` all'URI del log che hai utilizzato quando hai creato il cluster nel Passaggio 1. Il risultato sarà simile al seguente:

```
+-----+-----+
|          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	
+-----+	+-----+	+-----+

Cronologia dei rilasci di Phoenix

Nella seguente tabella sono elencati la versione di Phoenix inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente](#) per il rilascio in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni relative alla versione Phoenix

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.14.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.13.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
		library, hadoop-hdfs-namenode, hadoop-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-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

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.11.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.11.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.10.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.10.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.9.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.9.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.8.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.8.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.7.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.36.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.6.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.35.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.4.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.3.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.3.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.2.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.1.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.1.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.0.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-6.0.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.34.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.33.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.33.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.32.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.32.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.31.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.30.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.30.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.30.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.28.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.28.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.27.1	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.27.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.26.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.25.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.24.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.23.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.23.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.22.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.21.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.20.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.20.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.19.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.19.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.18.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.18.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.17.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.17.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.16.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.15.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.15.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.14.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.13.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.13.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.12.3	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.12.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.11.4	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.11.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.10.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.10.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.9.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.8.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.8.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.7.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.6.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.6.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.4	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.2	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
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

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-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-distcp, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-5.0.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-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-distcp, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.8.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.7.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di Phoenix	Componenti installati con Phoenix
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Apache Pig

Apache Pig è una libreria Apache open source eseguibile in Hadoop. Fornisce un linguaggio di script che puoi utilizzare per trasformare set di dati di grandi dimensioni senza dover scrivere codice complesso in un linguaggio di programmazione di livello inferiore come Java. La libreria accetta i comandi di tipo SQL scritti in un linguaggio denominato Pig Latin e li converte in processi Tez basati su grafi aciclici diretti (DAG) o programmi MapReduce. Pig utilizza dati strutturati e non strutturati in vari formati. Per ulteriori informazioni su Pig, visita la pagina Web all'indirizzo <http://pig.apache.org/>.

Puoi eseguire i comandi Pig in modalità interattiva o batch. Per utilizzare Pig in modalità interattiva, crea una connessione SSH al nodo master e invia i comandi utilizzando la shell Grunt. Per utilizzare Pig in modalità batch, scrivi script Pig, caricali in Amazon S3 e inviali come fasi del cluster. Per ulteriori informazioni sull'invio di lavoro a un cluster, consulta [Invio di lavoro a un cluster](#) nella Guida alla gestione di Amazon EMR.

Quando utilizzi Pig per scrivere l'output in una tabella HCatalog in Amazon S3, disabilita la scrittura diretta Amazon EMR impostando le proprietà `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` su `false`. Per ulteriori informazioni, consulta [Uso di HCatalog](#). In uno script Pig, puoi utilizzare i comandi `SET mapred.output.direct.NativeS3FileSystem false` e `SET mapred.output.direct.EmrFileSystem false`.

La seguente tabella indica la versione di Pig inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Pig.

Per la versione dei componenti installati con Pig in questo rilascio, consulta [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Pig per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-6.14.0	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-

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

La seguente tabella indica la versione di Pig inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Pig.

Per la versione dei componenti installati con Pig in questo rilascio, consulta [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Pig per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.36.1	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

Argomenti

- [Invio di lavoro Pig](#)
- [Chiamata di funzioni UDF \(definite dall'utente\) da Pig](#)

- [Cronologia dei rilasci di Pig](#)

Invio di lavoro Pig

Questa sezione illustra l'invio di lavoro Pig a un cluster Amazon EMR. Gli esempi che seguono generano un report contenente il numero totale di byte trasferiti, un elenco dei primi 50 indirizzi IP, un elenco dei primi 50 referrer esterni e i primi 50 termini di ricerca in Bing e Google. Lo script Pig si trova nel bucket Amazon S3 `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`. I dati di input si trovano nel bucket Amazon S3 `s3://elasticmapreduce/samples/pig-apache/input`. L'output viene salvato in un bucket Amazon S3.

Invio di lavoro Pig mediante la console di Amazon EMR

Questo esempio descrive come utilizzare la console di Amazon EMR per aggiungere una fase Pig a un cluster.

Per inviare una fase Pig

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegliere Creare un cluster per creare un cluster con Pig installato. Per i passaggi su come creare un cluster, consulta [Pianificare e configurare un cluster Amazon EMR](#).
3. Aprire una finestra del terminale e utilizzare SSH nel nodo principale del cluster seguendo la procedura descritta in [Connettersi al nodo principale tramite SSH](#). Una volta fatto questo, seguire i passaggi successivi.

```
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. Nella console, fare clic su Cluster List (Elenco cluster) e selezionare il nome del cluster creato.
5. Scorrere fino alla sezione Steps (Fasi) ed espanderla, quindi scegliere Add step (Aggiungi fase).
6. Nella finestra di dialogo Add Step (Aggiungi fase):
 - Per Step type (Tipo di fase), scegliere Pig program (programma Pig).
 - Per Name (Nome), accettare il nome predefinito (programma Pig) o digitare un nuovo nome.
 - Per Script S3 location (Percorso script S3), digitare il percorso dello script Pig. Ad esempio:
s3://elasticmapreduce/samples/pig-apache/do-reports2.pig.

- Per Script S3 location (Percorso script S3), digitare il percorso dei dati di input. Ad esempio: **s3://elasticmapreduce/samples/pig-apache/input**.
 - Per Output S3 location (Percorso S3 output), digita il nome del bucket di output Amazon S3.
 - Lasciare vuoto il campo Arguments (Argomenti).
 - Per Action on failure (Operazione in caso di errore), accettare l'opzione predefinita, ovvero Continue (Continua).
7. Scegliere Add (Aggiungi). La fase viene visualizzata nella console con lo stato Pending (In attesa).
 8. Durante l'esecuzione della fase, lo stato passa da Pending (In attesa) a Running (In esecuzione) a Completed (Completata). Per aggiornare lo stato, scegliere l'icona Refresh (Aggiorna) sopra la colonna Actions (Operazioni). Una volta completato il passaggio, controllare il bucket Amazon S3 per confermare che i file di output della fase Pig siano presenti.

Invio di lavoro Pig mediante la AWS CLI

Per inviare una fase Pig mediante l'AWS CLI

Quando si avvia un cluster mediante l'AWS CLI utilizzare il parametro `--applications` per installare Pig. Per inviare una fase Pig, utilizzare il parametro `--steps`.

1. Per avviare un cluster con Pig installato, digitare il comando seguente, sostituendo *myKey* e *DOC-EXAMPLE-BUCKET/* con il nome della coppia di chiavi EC2 e del bucket Amazon S3.

```
aws emr create-cluster \  
--name "Test cluster" \  
--log-uri s3://DOC-EXAMPLE-BUCKET/ \  
--release-label emr-5.36.1 \  
--applications Name=Pig \  
--use-default-roles \  
--ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge \  
--instance-count 3
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo master e le istanze rimanenti vengono avviate come nodi principali. Tutti i nodi utilizzano il tipo di istanza specificato nel comando.

Note

Se in precedenza non sono stati creati il ruolo del servizio EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

2. Per inviare una fase Pig, inserire il comando seguente, sostituendo *myClusterId* e *DOC-EXAMPLE-BUCKET* con l'ID cluster e il nome del bucket Amazon S3.

```
aws emr add-steps \  
--cluster-id myClusterId \  
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,Args=[-f,s3://  
elasticmapreduce/samples/pig-apache/do-reports2.pig,-p,INPUT=s3://elasticmapreduce/  
samples/pig-apache/input,-p,OUTPUT=s3://DOC-EXAMPLE-BUCKET/pig-apache/output]
```

Questo comando restituirà un ID fase che può essere utilizzato per controllare lo State della fase.

3. Esegui una query sullo stato della fase con il comando `describe-step`.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

Lo State della fase cambia da PENDING a RUNNING a COMPLETED durante la sua esecuzione. Una volta completato il passaggio, controllare il bucket Amazon S3 per confermare che i file di output della fase Pig siano presenti.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la [Guida di riferimento ai comandi della AWS CLI](#).

Chiamata di funzioni UDF (definite dall'utente) da Pig

Pig consente di chiamare funzioni definite dall'utente (UDF) da script Pig. Puoi eseguire questa operazione per implementare l'elaborazione personalizzata da utilizzare negli script Pig. I linguaggi attualmente supportati sono Java, Python/Jython e JavaScript (sebbene il supporto JavaScript sia ancora sperimentale).

Le seguenti sezioni descrivono come registrare le tue funzioni con Pig per chiamarle dalla shell Pig o dagli script Pig. Per ulteriori informazioni sull'utilizzo di funzioni UDF con Pig, consulta la [Documentazione Pig](#) relativa alla versione di Pig utilizzata.

Chiamata di file JAR da Pig

Puoi utilizzare file JAR personalizzati con Pig utilizzando il comando REGISTER nello script Pig. Il file JAR è un file system locale o remoto, ad esempio Amazon S3. Quando si esegue lo script Pig, Amazon EMR scarica il file JAR automaticamente nel nodo master e, successivamente, lo carica nella cache distribuita di Hadoop. In questo modo, il file JAR viene automaticamente utilizzato come necessario da tutte le istanze nel cluster.

Per utilizzare file JAR con Pig

1. Carica i file JAR personalizzati in Amazon S3.
2. Utilizza il comando REGISTER nello script Pig per specificare il bucket su Amazon S3 del file JAR personalizzato.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

Chiamata di script Python/Jython da Pig

Puoi registrare script Python con Pig e quindi chiamare funzioni in tali script dalla shell Pig o da uno script Pig. A questo scopo, devi specificare il percorso dello script con la parola chiave `register`.

Poiché Pig è scritto in Java, utilizza il modulo di gestione di script Jython per analizzare script Python. Per ulteriori informazioni su Jython, visita il sito Web all'indirizzo <http://www.jython.org/>.

Per chiamare uno script Python/Jython da Pig

1. Scrivi uno script Python e carica lo script in un percorso in Amazon S3. Questo deve essere un bucket appartenente allo stesso account che crea il cluster Pig o che dispone delle autorizzazioni impostate affinché l'account che ha creato il cluster possa accedervi. In questo esempio, lo script viene caricato in `s3://mybucket/pig/python`.
2. Avviare un cluster Pig. Se si accede a Pig dalla shell Grunt, eseguire un cluster interattivo. Se si eseguono comandi Pig da uno script, avviare un cluster Pig con script. Questo esempio avvia un cluster interattivo. Per ulteriori informazioni su come creare un cluster Pig, consulta [Invio di lavoro Pig](#).
3. Per un cluster interattivo, utilizzare SSH per connettersi al nodo master ed eseguire la shell Grunt. Per ulteriori informazioni, consulta [SSH nel nodo principale](#).
4. Eseguire la shell Grunt per Pig digitando `pig` nella riga di comando:

```
pig
```

5. Registra la libreria Jython e lo script Python con Pig utilizzando la parola chiave `register` nel prompt dei comandi Grunt, come illustrato nel comando seguente, in cui è necessario specificare il percorso dello script in Amazon S3:

```
grunt> register 'lib/jython.jar';  
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Caricare i dati di input. L'esempio seguente carica l'input da un percorso Amazon S3:

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as  
(line:chararray);
```

7. A questo punto, è possibile chiamare le funzioni nello script da Pig facendo riferimento alle stesse mediante `myfunctions`:

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

Cronologia dei rilasci di Pig

La seguente tabella indica la versione di Pig inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione

Versione componente per il rilascio in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione Pig

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-6.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-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,

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-6.7.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.36.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-6.4.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.3.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.33.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.30.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.30.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.29.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.22.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.21.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.17.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.16.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.16.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.15.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.12.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.12.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.3.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-5.2.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.1.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-resource-manager, 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-resource-manager, pig-client
emr-4.9.5	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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-4.8.0	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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-4.7.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.7.2	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-4.7.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
emr-4.7.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.6.0	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-4.5.0	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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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

Etichetta di rilascio di Amazon EMR	Version di Pig	Componenti installati con Pig
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 e Trino

Note

PrestoSQL è stato rinominato in Trino nel dicembre 2020. Amazon EMR versioni 6.4.0 e successive utilizzano il nome Trino, mentre le versioni precedenti utilizzano il nome PrestoSQL.

[Presto](#) è un motore di query SQL veloce, progettato per query analitiche interattive su dataset di grandi dimensioni da più origini. Per ulteriori informazioni, consulta il [sito Web Presto](#). Presto è incluso in Amazon EMR rilascio 5.0.0 e successivi. I rilasci precedenti includono Presto come applicazione sandbox (ambiente di sperimentazione). Per ulteriori informazioni, consulta [Versioni del rilascio di Amazon EMR 4.x](#). Amazon EMR versione 6.1.0 e successive supportano [Trino](#) (PrestoSQL) oltre a Presto. Per ulteriori informazioni, consulta [Installazione di PrestoDB e Trino](#).

La seguente tabella indica la versione di Presto inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Presto.

Per la versione dei componenti installati con Presto in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Presto per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.14.0	Presto 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-

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
		server, presto-coordinator, presto-worker

La seguente tabella indica la versione di Presto inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Presto.

Per la versione dei componenti installati con Presto in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Presto per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.36.1	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

La seguente tabella indica la versione di Trino (Presto SQL) inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Trino (Presto SQL).

Per la versione dei componenti installati con Trino (PrestoSQL) in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Trino (PrestoSQL) per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
emr-6.14.0	Trino (PrestoSQL) 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

Argomenti

- [Utilizzo di Presto con AWS Glue Data Catalog](#)
- [Uso di pushdown S3 Select con Presto per migliorare le prestazioni](#)
- [Aggiunta di connettori di database](#)
- [Utilizzo di SSL/TLS e configurazione di LDAPS con Presto on Amazon EMR](#)
- [Attivazione della modalità rigorosa Presto](#)
- [Gestione della perdita di istanze Spot in Presto](#)
- [Esecuzione con tolleranza ai guasti in Trino](#)
- [Utilizzo della scalabilità automatica Presto con disattivazione con tolleranza](#)
- [Considerazioni riguardanti Presto on Amazon EMR](#)
- [Cronologia delle versioni di Presto](#)

Utilizzo di Presto con AWS Glue Data Catalog

Utilizzando il rilascio di Amazon EMR versione 5.10.0 e successive, puoi specificare AWS Glue Data Catalog come metastore Hive predefinito per Presto. Si consiglia questa configurazione quando si

richiede un metastore persistente o un metastore condiviso da diversi cluster, servizi e applicazioni o Account AWS.

AWS Glue è un servizio di estrazione, trasformazione e caricamento (ETL) completamente gestito che consente di catalogare, filtrare e migliorare la qualità dei dati in modo semplice e conveniente, oltre a trasferirli in modo sicuro tra datastore. AWS Glue Data Catalog offre un repository di metadati unificato per varie origini dati e formati di dati, integrandosi con Amazon EMR così come con Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena e qualsiasi applicazione compatibile con il metastore Apache Hive. AWS I crawler di Glue possono dedurre automaticamente lo schema dai dati di origine in Amazon S3 e archiviare i metadati associati nel Data Catalog. Per ulteriori informazioni sul Data Catalog, consulta la sezione relativa al [Popolamento di AWS Glue Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

AWS Glue è soggetto a costi separati. Viene applicata una tariffa mensile per l'archiviazione e l'accesso ai metadati nel Data Catalog, una tariffa oraria fatturata al minuto per i processi ETL di AWS Glue e per il runtime del crawler e, infine, una tariffa oraria fatturata al minuto per ciascun endpoint di sviluppo di cui è stato effettuato il provisioning. Il Data Catalog consente di archiviare fino a un milione di oggetti senza costi aggiuntivi. Se intendi archiviare più di un milione di oggetti, ti verrà addebitato 1 USD per ogni 100.000 oggetti oltre il milione. Un oggetto nel Data Catalog può essere una tabella, una partizione o un database. Per ulteriori informazioni, consulta [Prezzi di Glue](#).

Important

Se hai creato tabelle utilizzando Amazon Athena o Amazon Redshift Spectrum prima del 14 agosto 2017, i database e le tabelle sono archiviati in un catalogo gestito da Athena, che è separato da AWS Glue Data Catalog. Per integrare Amazon EMR con queste tabelle, dovrai eseguire l'aggiornamento a AWS Glue Data Catalog. Per ulteriori informazioni, consulta [Aggiornamento a AWS Glue Data Catalog](#) nella Guida per l'utente di Amazon Athena.

Specifiche di AWS Glue Data Catalog come metastore

È possibile specificare AWS Glue Data Catalog come metastore utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Quando si utilizza la CLI o l'API, è possibile utilizzare la classificazione di configurazione per Presto per specificare un Data Catalog. Inoltre, con Amazon EMR 5.16.0 e versioni successive, è possibile utilizzare la classificazione di configurazione per specificare un Data Catalog in un Account AWS diverso. Quando utilizzi la console, puoi specificare il Data Catalog utilizzando Advanced Options (Opzioni avanzate) o Quick Options (Opzioni rapide).

New console

Specifica di AWS Glue Data Catalog come metastore Hive con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Application bundle (Bundle di applicazioni), scegli Presto.
4. In Impostazioni del Catalogo dati AWS Glue, seleziona la casella di controllo Utilizza per i metadati delle tabelle Presto.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di AWS Glue Data Catalog come metastore Presto predefinito con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Software Configuration (Configurazione software) scegliere una Release (Rilascio) di emr-5.10-0 o versione successiva e selezionare Presto.
4. Selezionare Use for Presto table metadata (Utilizza per i metadati delle tabelle Presto), scegliere Next (Successivo), quindi completare altre impostazioni per il cluster come appropriato per l'applicazione.

CLI

Specifica di AWS Glue Data Catalog come metastore Hive di default con la AWS CLI

Per esempi su come specificare le seguenti classificazioni di configurazione al momento della creazione di un cluster, consulta [Configurazione delle applicazioni](#).

Amazon EMR 5.16.0 e versioni successive

- Imposta la proprietà `hive.metastore` su `glue`, come mostrato nell'esempio JSON seguente.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Per specificare un Catalogo di dati in un Account AWS differente, aggiungi la proprietà `hive.metastore.glue.catalogid` come illustrato nel seguente esempio JSON. Sostituisci *acct-id* con l'account AWS di Data Catalog. Non puoi utilizzare un Catalogo di dati in un altro Account AWS se utilizzi Amazon EMR versione 5.15.0 e precedenti.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

Amazon EMR 5.10.0 fino a 5.15.0

Impostare la proprietà `hive.metastore.glue.datacatalog.enabled` su `true`, come mostrato nell'esempio JSON seguente:

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

Amazon EMR 6.1.0 e versioni successive utilizzando PrestoSQL (Trino)

A partire da EMR versione 6.1.0, PrestoSQL supporta anche Glue come metastore Hive predefinito. Utilizza la classificazione della configurazione `prestoql-connector-hive` e imposta la proprietà `hive.metastore` su `glue` come mostrato nel seguente esempio di JSON.

Amazon EMR versione 6.4.0 e successive utilizza il nuovo nome Trino anziché PrestoSQL. Se usi Trino, sostituisci `prestoql-connector-hive` nella seguente classificazione di configurazione con `trino-connector-hive`.

```
[
  {
    "Classification": "prestoql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Per cambiare metastore di un cluster di lunga esecuzione, è possibile impostare manualmente questi valori come appropriati per la tua versione di release collegandoti al nodo master, modificando i valori di proprietà direttamente nel file `/etc/presto/conf/catalog/hive.properties` e riavviando il server Presto (`sudo restart presto-server`). Se utilizzi questo metodo con Amazon EMR 5.15.0 e versioni precedenti, assicurati che `hive.table-statistics-enabled` sia impostato su `false`. Questa impostazione non è necessaria quando si usano le versioni della release 5.16.0 e successive; tuttavia, le statistiche di tabella e di partizione non sono supportate.

Autorizzazioni IAM

Il profilo dell'istanza EC2 per un cluster deve disporre di autorizzazioni IAM per le operazioni AWS Glue. Inoltre, se abiliti la crittografia per oggetti AWS Glue Data Catalog, al ruolo deve essere consentito anche crittografare, decrittare e generare la AWS KMS key utilizzata per la crittografia dei dati.

Autorizzazioni per operazioni AWS Glue

Se utilizzi il profilo dell'istanza EC2 di default per Amazon EMR, non è richiesta alcuna operazione. La policy gestita `AmazonElasticMapReduceforEC2Role` collegata al `EMR_EC2_DefaultRole` consente tutte le operazioni AWS Glue necessarie. Tuttavia, se specifichi autorizzazioni e un profilo dell'istanza EC2 personalizzati, devi configurare le operazioni AWS Glue appropriate. Utilizza la policy `AmazonElasticMapReduceforEC2Role` gestita come punto di partenza. Per ulteriori informazioni, consulta [Ruolo di servizio per le istanze EC2 del cluster \(profilo dell'istanza EC2\)](#) nella Guida alla gestione di Amazon EMR.

Autorizzazioni per crittografare e decrittografare AWS Glue Data Catalog

Il tuo profilo dell'istanza ha bisogno dell'autorizzazione per crittografare e decrittare i dati utilizzando la tua chiave. Non è necessario configurare queste autorizzazioni se si applicano entrambe le seguenti istruzioni:

- Per abilitare la crittografia per oggetti AWS Glue Data Catalog utilizzare chiavi gestite per AWS Glue.
- Si utilizza un cluster che è nello stesso Account AWS come il AWS Glue Data Catalog.

In caso contrario, devi aggiungere la istruzione seguente alla policy delle autorizzazioni collegata al profilo dell'istanza EC2.

```
[
  {
    "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"
      }
    ]
  }
]
```

Per ulteriori informazioni sulla crittografia di AWS Glue Data Catalog, consulta la sezione [Crittografia del proprio Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

Autorizzazioni basate su risorse

Se utilizzi AWS Glue in combinazione con Hive, Spark o Presto in Amazon EMR, AWS Glue supporta policy basate su risorse per controllare l'accesso alle risorse del Data Catalog. Queste risorse includono tabelle, database, connessioni e funzioni definite dall'utente. Per ulteriori informazioni, consulta [Policy sulle risorse AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.

Quando utilizzi policy basate su risorse per limitare l'accesso ad AWS Glue da Amazon EMR, l'entità principale specificata nella policy delle autorizzazioni deve essere il ruolo ARN associato al profilo di istanza EC2 specificato al momento della creazione di un cluster. Ad esempio, per una policy basata sulle risorse collegata a un catalogo, è possibile specificare il ruolo ARN per il ruolo di servizio predefinito per le istanze del cluster EC2, *EMR_EC2_DefaultRole*, come `Principal` utilizzando il formato mostrato nell'esempio seguente:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

L'*acct-id* può essere diverso dall'ID dell'account AWS Glue. Ciò consente l'accesso da cluster EMR in account diversi. Puoi specificare più entità principali, ognuna da un account diverso.

Considerazioni sull'utilizzo di AWS Glue Data Catalog

Considera i seguenti elementi durante l'utilizzo di AWS Glue Data Catalog come metastore con Presto:

- La ridenominazione delle tabelle in AWS Glue non è supportata.
- Quando si crea una tabella Hive senza specificare una `LOCATION`, i dati nella tabella vengono archiviati nella posizione specificata dalla proprietà `hive.metastore.warehouse.dir`. Per impostazione predefinita, questa è una posizione in HDFS. Se un altro cluster necessita di accedere alla tabella, non ci riuscirà a meno che non disponga di autorizzazioni adeguate per il cluster che ha creato la tabella. Inoltre, poiché l'archiviazione HDFS è transitoria, se il cluster viene terminato, i dati della tabella vanno persi e la tabella deve essere ricreata. Ti consigliamo di specificare una `LOCATION` in Amazon S3 quando crei una tabella Hive utilizzando AWS Glue. In alternativa, puoi utilizzare la classificazione di configurazione `hive-site` per specificare una posizione in Amazon S3 per `hive.metastore.warehouse.dir`, che si applica a tutte le tabelle Hive. Se una tabella viene creata in una posizione HDFS e il cluster che l'ha creata è ancora in

esecuzione, puoi aggiornare la posizione della tabella ad Amazon S3 in AWS Glue. Per ulteriori informazioni, consulta [Utilizzo di tabelle sulla console AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.

- I valori delle partizioni contenenti virgolette e apostrofi non sono supportati, ad esempio `PARTITION (owner="Doe 's")..`
- Le [statistiche di colonna](#) sono supportate per emr-5.31.0 e versioni successive.
- L'utilizzo dell'[autorizzazione Hive](#) non è supportato. In alternativa, puoi prendere in considerazione l'utilizzo di [Policy basate sulle risorse AWS Glue](#). Per ulteriori informazioni, consulta [Utilizzo delle policy basate sulle risorse per l'accesso di Amazon EMR a AWS Glue Data Catalog](#).

Uso di pushdown S3 Select con Presto per migliorare le prestazioni

Con Amazon EMR 5.18.0 e versioni successive, puoi utilizzare Pushdown [S3 Select](#) con Presto on Amazon EMR. Questa funzione consente a Presto di "eseguire il pushdown" dell'attività di calcolo delle operazioni di proiezione (ad esempio, SELECT) e le operazioni di predicato (ad esempio, WHERE) ad Amazon S3. Ciò consente alle query di recuperare solo i dati richiesti da Amazon S3, cosa che può migliorare le prestazioni e ridurre la quantità di dati trasferiti tra Amazon EMR e Amazon S3 in alcune applicazioni.

Pushdown S3 Select è adatto alla mia applicazione?

È consigliabile eseguire analisi comparative delle applicazioni con e senza pushdown S3 Select per vedere se l'utilizzo di tale funzionalità può risultare idoneo per l'applicazione.

Per determinare se l'applicazione è idonea per l'uso di S3 Select, utilizza le seguenti linee guida:

- La query filtra più di metà del set di dati originale.
- I predicati del filtro della query utilizzano colonne che dispongono di un tipo di dati supportato da Presto e S3 Select. I tipi di dati timestamp, reali e doppi non sono supportati da pushdown S3 Select. Ti consigliamo di usare il tipo di dati decimale per i dati numerici. Per ulteriori informazioni sui tipi di dati supportati per S3 Select, consulta [Tipi di dati](#) nella Guida per l'utente di Amazon Simple Storage Service.
- La connessione di rete tra Amazon S3 e il cluster Amazon EMR ha una buona velocità di trasferimento e larghezza di banda disponibile. Amazon S3 non comprime le risposte HTTP, pertanto è probabile che le dimensioni della risposta aumentino per file di input compressi.

Considerazioni e limitazioni

- Sono supportati solo gli oggetti archiviati in formato CSV. Gli oggetti possono essere decompressi o facoltativamente compressi con gzip o bzip2.
- La proprietà `AllowQuotedRecordDelimiters` non è supportata. Se questa proprietà è specificata, la query ha esito negativo.
- La crittografia lato server con chiavi di crittografia fornite dal cliente (SSE-C) e la crittografia lato client di Amazon S3 non sono supportate.
- Pushdown S3 Select non è un'alternativa per l'uso di formati di file compressi o basati su colonne, ad esempio ORC o Parquet.

Abilitare Pushdown S3 Select con PrestoDB o Trino

Per abilitare pushdown S3 Select con PrestoDB su Amazon EMR, utilizza la classificazione di configurazione `presto-connector-hive` per impostare `hive.s3select-pushdown.enabled` su `true` come illustrato nell'esempio seguente. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#). Imposta anche il valore `hive.s3select-pushdown.max-connections`. Per la maggior parte delle applicazioni, l'impostazione di **500** dovrebbe essere adeguata. Per ulteriori informazioni, consulta [Comprensione e ottimizzazione di `hive.s3select-pushdown.max-connections`](#) di seguito.

Per PrestoSQL su EMR versioni 6.1.0 - 6.3.0, sostituisci *`presto-connector-hive`* nell'esempio sottostante con `prestoql-connector-hive`.

Amazon EMR versione 6.4.0 e successive utilizza il nuovo nome Trino anziché PrestoSQL. Se usi Trino, sostituisci *`presto-connector-hive`* nell'esempio sottostante con `trino-connector-hive`

```
[
  {
    "classification": "presto-connector-hive",
    "properties": {
      "hive.s3select-pushdown.enabled": "true",
      "hive.s3select-pushdown.max-connections": "500"
    }
  }
]
```

Comprensione e ottimizzazione di `hive.s3select-pushdown.max-connections`

Per impostazione predefinita, Presto usa EMRFS come file system. L'impostazione `fs.s3.maxConnections` nella classificazione di configurazione `emrfs-site` specifica il numero massimo consentito di connessioni client ad Amazon S3 tramite EMRFS per Presto. Per impostazione predefinita, questo è 500. Pushdown S3 Select ignora EMRFS durante l'accesso ad Amazon S3 per le operazioni di predicato. In questo caso, il valore di `hive.s3select-pushdown.max-connections` determina il numero massimo di connessioni client consentite per tali operazioni dai nodi di lavoro. Tuttavia, tutte le richieste ad Amazon S3 avviate da Presto che non vengono spinte verso il basso, ad esempio le operazioni GET, continuano ad essere regolate dal valore di `fs.s3.maxConnections`.

Se l'applicazione riceve l'errore di timeout in attesa di connessione dal pool, aumentare il valore di `hive.s3select-pushdown.max-connections` e di `fs.s3.maxConnections`.

Aggiunta di connettori di database

Puoi usare le classificazioni di configurazione per configurare le proprietà del connettore JDBC al momento della creazione di un cluster. Le classificazioni di configurazione iniziano con `presto-connector`, ad esempio, `presto-connector-postgresql`. Le classificazioni di configurazione disponibili dipendono dalla versione del rilascio Amazon EMR. Per le classificazioni di configurazione disponibili con la versione del rilascio più recente, consulta [the section called “Classificazioni di configurazione”](#) per Amazon EMR 5.36.1. Se stai utilizzando un'altra versione di Amazon EMR, consulta [Versioni del rilascio di Amazon EMR 5.x](#) per le classificazioni di configurazione. Per ulteriori informazioni sulle proprietà che possono essere configurate con ciascun connettore, consulta <https://prestodb.io/docs/current/connector.html>.

Example Configurazione di un cluster con il connettore PostgreSQL JDBC

Per avviare un cluster con il connettore PostgreSQL installato e configurato, crea innanzitutto un file JSON che specifica la classificazione della configurazione, ad esempio `myConfig.json`, con il seguente contenuto e salvalo in locale.

Sostituisci le proprietà di connessione appropriate per la configurazione come illustrato nell'argomento [PostgreSQL connector \(Connettore PostgreSQL\)](#) nella documentazione di Presto.

```
[  
  {
```

```
"Classification": "presto-connector-postgresql",
"Properties": {
  "connection-url": "jdbc:postgresql://example.net:5432/database",
  "connection-user": "MYUSER",
  "connection-password": "MYPASS"
},
"Configurations": []
}
]
```

Quando si crea il cluster, fare riferimento al percorso del file JSON utilizzando l'opzione `--configurations` come nell'esempio seguente, dove `myConfig.json` si trova nella stessa directory in cui si esegue il comando:

```
aws emr create-cluster --name PrestoConnector --release-label emr-5.36.1 --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
```

Utilizzo di SSL/TLS e configurazione di LDAPS con Presto on Amazon EMR

Con il rilascio di Amazon EMR versione 5.6.0 e successive, è possibile abilitare SSL/TLS per facilitare la [comunicazione interna sicura](#) tra i nodi Presto. A questo scopo, impostare una configurazione di protezione per la crittografia dei dati in transito. Per ulteriori informazioni, consulta [Opzioni di crittografia](#) e [Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster](#) nella Guida alla gestione di Amazon EMR.

Quando si utilizza una configurazione di sicurezza con la crittografia dei dati in transito, Amazon EMR esegue le seguenti operazioni per Presto:

- Distribuisce gli artefatti per la crittografia o certificati specificati per la crittografia in transito in tutto il cluster Presto. Per ulteriori informazioni, consulta [Fornitura di certificati per la crittografia dei dati in transito](#).
- Imposta le seguenti proprietà utilizzando la classificazione di configurazione `presto-config` che corrisponde al file `config.properties` per Presto:

- Imposta `http-server.http.enabled` su `false` per tutti i nodi, il che disabilita HTTP a favore di HTTPS. In questo caso è necessario fornire certificati che funzionino per DNS pubblico e privato quando si imposta la configurazione di protezione per la crittografia in transito. Un modo per farlo consiste nell'utilizzare i certificati SAN (Subject Alternative Name) che supportano più domini.
- Imposta i valori `http-server.https.*`. Per dettagli sulla configurazione, consulta [Autenticazione LDAP](#) nella documentazione di Presto.
- Per PrestoSQL (Trino) su EMR versione 6.1.0 e successive, Amazon EMR configura automaticamente una chiave segreta condivisa per una comunicazione interna sicura tra i nodi del cluster. Non è necessario eseguire alcuna configurazione aggiuntiva per abilitare questa funzionalità di protezione ed è possibile ignorare la configurazione con la propria chiave segreta. Per informazioni sull'autenticazione interna di Trino, consulta [Documentazione Trino 353: comunicazione interna sicura](#).

Inoltre, con il rilascio di Amazon EMR versione 5.10.0 e successive, è possibile impostare l'[Autenticazione LDAP](#) per le connessioni client al coordinatore Presto tramite HTTPS. Questa impostazione utilizza LDAP sicuro (LDAPS). TLS deve essere abilitato sul server LDAP e il cluster Presto deve utilizzare una configurazione di sicurezza con la crittografia dei dati in transito abilitata. È richiesta una configurazione aggiuntiva. Le opzioni di configurazione sono diverse a seconda della versione del rilascio di Amazon EMR utilizzata. Per ulteriori informazioni, consulta [Utilizzo dell'autenticazione LDAP con Presto on Amazon EMR](#).

Presto on Amazon EMR utilizza la porta 8446 per HTTPS interno per impostazione predefinita. La porta utilizzata per la comunicazione interna deve essere la stessa porta utilizzata per l'accesso del client HTTPS al sistema di coordinamento Presto. La proprietà `http-server.https.port` nella classificazione di configurazione `presto-config` specifica la porta.

Utilizzo dell'autenticazione LDAP con Presto on Amazon EMR

Seguire le procedure di questa sezione per configurare LDAP. Vedi ogni passo per esempi e link per ulteriori informazioni.

Fasi per configurare l'autenticazione LDAP

- [Fase 1: è possibile ottenere informazioni sui server LDAP e copiare il certificato server su Amazon S3](#)
- [Fase 2: impostazione di una configurazione di sicurezza](#)

- [Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP](#)
- [Fase 4: creare lo script per copiare il certificato del server LDAP e caricarlo su Amazon S3](#)
- [Fase 5: creazione del cluster](#)

Fase 1: è possibile ottenere informazioni sui server LDAP e copiare il certificato server su Amazon S3

Sarà necessario disporre delle informazioni e degli elementi nella sezione seguente dal server LDAP per configurare l'autenticazione LDAP.

L'indirizzo IP o il nome host del server LDAP

Il coordinatore Presto sul nodo master Amazon EMR deve essere in grado di raggiungere il server LDAP all'indirizzo IP o al nome host specificato. Per impostazione predefinita, Presto comunica con il server LDAP utilizzando LDAPS tramite la porta 636. Se l'implementazione LDAP richiede una porta personalizzata, è possibile specificarla utilizzando la proprietà `ldap.url` con Amazon EMR 5.16.0 o versioni successive o utilizzando `authentication.ldap.url` con versioni precedenti. Sostituire la porta personalizzati per 636 come mostrato negli esempi di classificazione della configurazione `presto-config` in [Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP](#). Verificare che i firewall e i gruppi di sicurezza consentano il traffico in entrata e in uscita sulla porta 636 (o la porta personalizzata) e anche sulla porta 8446 (o la porta personalizzata), che viene utilizzata per le comunicazioni interne del cluster.

Il certificato del server LDAP

È necessario caricare il file del certificato in un percorso sicuro in Amazon S3. Per ulteriori informazioni, consulta [Come caricare file e cartelle in un bucket S3](#) nella Guida per l'utente di Amazon Simple Storage Service. È possibile creare un'operazione di bootstrap che copi questo certificato da Amazon S3 su ogni nodo nel cluster quando il cluster viene avviato. Nello stato [Fase 4: creare lo script per copiare il certificato del server LDAP e caricarlo su Amazon S3](#). Il certificato di esempio è `s3://MyBucket/ldap_server.crt`.

Le impostazioni del server LDAP per binding anonimo

Se il binding anonimo è disattivato su PrestoDB, sono necessari l'ID utente (UID) e la password di un account con le autorizzazioni per vincolarsi al server LDAP affinché il server PrestoDB sia in grado di stabilire una connessione. È possibile specificare l'UID e la password utilizzando le proprietà `internal-communication.authentication.ldap.user` e `internal-`

`communication.authentication.ldap.password` e `presto-config` nella classificazione delle configurazioni. Amazon EMR 5.10.0 non supporta queste impostazioni, per cui il binding anonimo deve essere supportato sul server LDAP quando si utilizza questa versione del rilascio.

Occorre notare che Trino non richiede la configurazione dell'associazione anonima.

Per ottenere lo stato del binding anonimo sul server LDAP

- Utilizza il comando [ldapwhoami](#) di un client Linux, come mostrato nel seguente esempio:

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```

Se il binding anonimo non è consentito, il comando restituisce quanto segue:

```
ldap_bind: Inappropriate authentication (48)
additional info: anonymous bind disallowed
```

Per verificare che un account abbia le autorizzazioni per un server LDAP che utilizza l'autenticazione semplice

- Utilizza il comando [ldapwhoami](#) di un client Linux, come mostrato nel seguente esempio. L'esempio utilizza un utente fittizio, *presto*, archiviato in un server Open LDAP in esecuzione su un'istanza EC2 con il nome host fittizio *ip-xxx-xxx-xxx-xxx.ec2.internal*. L'utente è associato all'unità organizzativa (OU) *admins* e alla password *123456*:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal
```

Se l'account è valido e dispone di autorizzazioni appropriate, il comando restituisce:

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

Le configurazioni di esempio in [Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP](#) includono questo account per chiarezza, ad eccezione dell'esempio 5.10.0, dove non è supportato. Se il server LDAP usa un binding anonimo, rimuovere le coppie nome/valore `internal-communication.ldap.user` e `internal-communication.ldap.password`.

Il nome distinto LDAP (DN) per gli utenti Presto

Quando si specifica la configurazione LDAP per Presto, è necessario specificare un modello bind che comprende `{USER}` insieme a un'unità organizzativa (OU) e altri componenti di dominio aggiuntivi (DC). Presto sostituisce `{USER}` con l'ID utente (UID) attuale di ciascun utente durante l'autenticazione delle password per correggere il nome distinto (DN) che questo modello bind specifica. È necessario avere gli UO a cui appartengono gli utenti idonei e i loro DC. Ad esempio, per consentire agli utenti dell'UO `admins` nel dominio `corp.example.com` di autenticarsi per Presto, è necessario specificare `{USER},ou=admins,dc=corp,dc=example,dc=com` come modello bind dell'utente.

Note

Quando si utilizza AWS CloudFormation, è necessario utilizzare la funzione `Fn::Sub` per sostituire `{USER}` con l'ID utente (UID) effettivo. Per ulteriori informazioni, consulta il argomento [Fn::Sub](#) nella Guida per l'utente di AWS CloudFormation.

Con Amazon EMR 5.10.0, è possibile specificare un solo modello di questo genere. Utilizzando Amazon EMR 5.11.0 o versioni successive, è possibile specificare più modelli separati da due punti (:). Gli utenti che tentano di effettuare l'autenticazione su Presto sono paragonati al primo modello, quindi al secondo e così via. Per vedere un esempio, consulta [Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP](#).

Fase 2: impostazione di una configurazione di sicurezza

Creare una configurazione di sicurezza con la crittografia dei dati in transito abilitata. Per ulteriori informazioni, consulta [Creazione di una configurazione di sicurezza](#) nella Guida per la gestione di Amazon EMR. Gli artefatti della crittografia forniti durante la configurazione della crittografia dei dati in transito vengono utilizzati per crittografare le comunicazioni interne tra i nodi Presto. Per ulteriori informazioni, consulta [Fornitura di certificati per la crittografia dei dati in transito](#). Il certificato del server LDAP viene utilizzato per autenticare le connessioni client al server di Presto.

Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP

È possibile utilizzare la classificazione delle configurazioni `presto-config` per impostare le proprietà per LDAP. Il formato e il contenuto di `presto-config` sono leggermente diversi a seconda della versione di Amazon EMR e dell'installazione di Presto (PrestoDB o Trino). Più avanti in questa

sezione verranno forniti alcuni esempi di differenze di configurazione. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

La seguente procedura presuppone che i dati JSON vengano salvati in un file, *MyPrestoConfig.json*. Se si utilizza la console, carica il file in un percorso sicuro in Amazon S3 in modo che sia possibile farvi riferimento al momento della creazione del cluster. Se si utilizza AWS CLI, è possibile fare riferimento al file in locale.

Example Amazon EMR 6.1.0 e versioni successive con PrestoSQL (Trino)

L'esempio seguente usa il nome host LDAP da [Fase 1: è possibile ottenere informazioni sui server LDAP e copiare il certificato server su Amazon S3](#) per eseguire l'autenticazione sul server LDAP per il binding. Vengono specificati due modelli di binding degli utenti, che indicano che gli utenti all'interno dell'UO admins e dell'UO datascientists sul server LDAP sono idonei per l'autenticazione al server Trino come utenti. I modelli di binding sono separati da due punti (:).

Amazon EMR versione 6.4.0 e successive utilizza il nuovo nome Trino anziché PrestoSQL. Se usi Trino, sostituisci *prestoql-config* nella seguente classificazione di configurazione con *trino-config* e *prestoql-password-authenticator* con *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 e versioni successive

L'esempio seguente usa la password, l'ID utente e il nome host LDAP da [Fase 1: è possibile ottenere informazioni sui server LDAP e copiare il certificato server su Amazon S3](#) per eseguire

l'autenticazione sul server LDAP per il binding. Vengono specificati due modelli di binding degli utenti, che indicano che gli utenti all'interno dell'UO `admins` e dell'UO `datascientists` sul server LDAP sono idonei per l'autenticazione al server Presto come utenti. I modelli di binding sono separati da due punti (:).

```
[{
  "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 fino a 5.15.0

Il formato della classificazione di configurazione `presto-config` è leggermente diverso per queste versioni di release. L'esempio seguente specifica gli stessi parametri del precedente.

```
[{
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "LDAP",
    "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
    "authentication.ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
    "internal-communication.authentication.ldap.user": "presto",
    "internal-communication.authentication.ldap.password": "123456"
  }
}]
```

Example Amazon EMR 5.10.0

Amazon EMR 5.10.0 supporta solo il binding anonimo, per cui tali voci sono omesse. Inoltre, può essere specificato solo un singolo modello di binding.

```
[{
  "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"
  }
}]
```

Fase 4: creare lo script per copiare il certificato del server LDAP e caricarlo su Amazon S3

Crea uno script che copia il file di certificato su ogni nodo nel cluster e lo aggiunge al keystore. Crea lo script utilizzando un editor di testo, salvalo e quindi caricalo su Amazon S3. In [Fase 5: creazione del cluster](#), il file di script viene mostrato come `s3://MyBucket/LoadLDAPCert.sh`.

Lo script seguente di esempio utilizza la password keystore predefinita, `changeit`. È consigliabile connettersi al nodo master dopo aver creato il cluster e modificare la password del keystore utilizzando il comando `keytool`.

```
#!/bin/bash
aws s3 cp s3://MyBucket/ldap_server.crt .
sudo keytool -import -keystore /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/lib/security/
cacerts -trustcacerts -alias ldap_server -file ./ldap_server.crt -storepass changeit -
noprompt
```

Fase 5: creazione del cluster

Quando crei il cluster, devi specificare Presto e altre applicazioni che desideri che Amazon EMR installi. I seguenti esempi fanno anche riferimento alle proprietà di classificazione della configurazione all'interno di un JSON, ma è anche possibile specificare la classificazione di configurazione inline.

Creazione di un cluster Presto con autenticazione LDAP utilizzando la console di Amazon EMR

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Scegli Presto insieme ad altre applicazioni che Amazon EMR deve installare e in Software Configuration (Configurazione software), seleziona il Release (Rilascio) di Amazon EMR da utilizzare. L'autenticazione LDAP è supportata solo con Amazon EMR 5.10.0 e versioni successive.
4. In Edit software settings (Modifica le impostazioni software), scegli Load JSON from S3 (Carica JSON da S3), inserisci il percorso in Amazon S3 del file di configurazione JSON creato in [Fase 3: crea una configurazione JSON con le proprietà Presto per LDAP](#), quindi scegli Next (Successivo).
5. Configura la rete e l'hardware del cluster, quindi scegli Next (Successivo).
6. Seleziona Bootstrap actions (Operazioni di bootstrap). Per Add bootstrap action (Aggiungi operazione di bootstrap), seleziona Custom action (Operazione personalizzata), quindi scegli Configure and add (Configura e aggiungi).
7. Inserisci un Name (Nome) per l'operazione di bootstrap, immetti la Script location (Posizione script) creato in [Fase 4: creare lo script per copiare il certificato del server LDAP e caricarlo su Amazon S3](#), ad esempio s3://MyBucket/LoadLDAPCert.sh, quindi scegli Add (Aggiungi).
8. In General Options (Opzioni generali), Tag e Additional Options (Opzioni aggiuntive), scegli le impostazioni appropriate per la tua applicazione, quindi scegli Next (Successivo).
9. Scegli Authentication and encryption (Autenticazione e crittografia), quindi seleziona la Security configuration (Configurazione di sicurezza) creata in [Fase 2: impostazione di una configurazione di sicurezza](#).
10. Scegliere altre opzioni di sicurezza come appropriato per l'applicazione, quindi scegliere Create Cluster (Crea cluster).

Per creare un cluster Presto con l'autenticazione LDAP usando l'AWS CLI

- Utilizza il comando `aws emr create-cluster`. Come minimo, specificare l'applicazione Presto e la classificazione della configurazione Presto, lo script di bootstrap e la configurazione di sicurezza creata nei passaggi precedenti. L'esempio seguente fa riferimento al file i configurazione come a un file JSON salvato nella stessa directory in cui viene eseguito il

comando: Lo script di bootstrap, invece, deve essere salvato in Amazon S3. Nell'esempio seguente viene utilizzato `s3://MyBucket/LoadLDAPCert.sh`.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (`^`).

```
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
```

Attivazione della modalità rigorosa Presto

In determinate situazioni, le query a esecuzione prolungata possono comportare costi elevati e far sì che Amazon EMR utilizzi più risorse del cluster. Questo sottrae risorse agli altri carichi di lavoro del cluster. Con Amazon EMR versione 6.8 e successive, puoi utilizzare una funzionalità in modalità rigorosa che rifiuta o ti avvisa dei seguenti tipi di query a lunga durata:

- Query senza predicati sulle colonne con partizioni che generano scansioni di tabelle di grandi quantità di dati
- Query con cross join tra due tabelle di grandi dimensioni
- Query che ordinano un numero elevato di righe senza limiti

Dopo che Presto ottimizza completamente il piano di query, viene eseguita la modalità rigorosa. Per utilizzare e personalizzare la modalità rigorosa in base alle tue esigenze di esecuzione delle query, puoi configurare Presto nei seguenti modi.

Configurazioni Presto per la modalità rigorosa

Impostazione	Descrizione	Default
<code>strict-mode-enabled</code>	Attiva e disattiva la modalità rigorosa. Il valore <code>true</code> indica che la modalità rigorosa è attiva.	<code>false</code>
<code>strict-mode-fail-query</code>	Rifiuta le query se la modalità rigorosa rileva probabili query a lunga durata. Se <code>false</code> , Amazon EMR genera solo un avviso.	<code>false</code>
<code>strict-mode-restrictions</code>	Specifica le limitazioni da applicare quando è abilitata la modalità rigorosa. La modalità rigorosa supporta le seguenti limitazioni: <code>MANDATORY_PARTITION_PREDICATE</code> , <code>DISALLOW_CROSS_JOIN</code> e <code>LIMITED_SORT</code> .	<code>MANDATORY_PARTITION_PREDICATE,DISALLOW_CROSS_JOIN,LIMITED_SORT</code>

Per sperimentare la modalità rigorosa, puoi sovrascrivere queste configurazioni o impostarle come proprietà di sessione quando usi il client Presto.

Impostazione della configurazione al momento della creazione del cluster con la AWS Management Console

1. Scegli **Create cluster** (Crea cluster) e seleziona Amazon EMR versione 6.8.0 e Presto o Trino. Per ulteriori informazioni, consulta [Installazione di PrestoDB e Trino](#).
2. Specifica direttamente le proprietà di configurazione per la modalità rigorosa o carica un file JSON su Amazon S3. Facoltativamente, seleziona il per il tuo metastore. Specifica il tuo VPC, le sottoreti, le azioni di bootstrap, la coppia di chiavi e il gruppo di sicurezza. Per creare il cluster, scegli **Create cluster** (Crea cluster).
3. Accedi al nodo primario del cluster ed esegui `presto-cli` o `trino-cli`.

4. Invia le tue query. La modalità rigorosa convalida ogni query e determina se è di lunga durata. A seconda dell'impostazione di `strict-mode-fail-query`, Amazon EMR rifiuta la query o genera un avviso.
5. Quando hai finito con le query, termina il cluster ed elimina le tue risorse.

Impostazione della configurazione su un cluster in esecuzione con la AWS CLI

1. Accedi al nodo primario del cluster con la AWS CLI ed esegui `presto-cli` o `trino-cli`.
2. Emetti i seguenti comandi con i valori desiderati.

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

Considerazioni

Quando utilizzi la modalità rigorosa, considera quanto riportato di seguito:

- In alcuni casi, la modalità rigorosa può rifiutare le query di breve durata che non utilizzano numerose risorse. Ad esempio, le query su tabelle di piccole dimensioni non applicano filtri dinamici né sostituiscono i join interni con unioni incrociate. Ciò può portare la query a utilizzare il predicato di partizione obbligatorio o a impedire il cross join. Quando ciò accade, la modalità rigorosa rifiuta la query.
- Il controllo in modalità rigorosa viene applicato solo ai tipi di query `SELECT`, `INSERT`, `CREATE TABLE AS SELECT` e `EXPLAIN`.
- Puoi utilizzare la limitazione sul predicato di partizione obbligatorio solo con il connettore Hive.

Gestione della perdita di istanze Spot in Presto

Con le istanze Spot in Amazon EMR, puoi eseguire carichi di lavoro di Big Data su capacità Amazon EC2 di riserva a un costo ridotto. In cambio di costi inferiori, Amazon EC2 può interrompere le istanze Spot con una notifica di due minuti. Quando termini un nodo, Presto può impiegare fino a 10 minuti prima di restituire un errore. Ciò causa inutili ritardi nelle segnalazioni di errori e possibili nuovi tentativi. La terminazione rapida è una funzionalità che consente di controllare il modo in cui Presto gestisce i nodi terminati.

Il processo del coordinatore Presto è quello di tenere traccia di tutti i nodi worker con sondaggi regolari del loro status. Senza una terminazione rapida, il coordinatore non consulta lo YARN NodeManager per lo stato di ogni nodo. Ciò può comportare un lungo ciclo di nuovi tentativi prima che la query abbia esito negativo. Con una terminazione rapida, il coordinatore Presto consulta lo stato del nodo nel NodeManager non appena il sondaggio non riesce a raggiungere l'host. Se NodeManager mostra che il nodo è inattivo, Presto abbandona ulteriori tentativi, fallisce la query e restituisce un errore `NODE_DECOMMISSIONED`.

Il seguente set di parametri di configurazione consente di controllare e personalizzare il comportamento di Presto in caso di terminazione del nodo.

Configurazioni Presto per la gestione degli errori dei nodi

Impostazione	Descrizione	Default
<code>query.remote-task.max-backoff-duration</code>	Il periodo di tempo in cui il coordinatore continua a provare a recuperare lo stato dell'attività remota dai nodi worker.	10 minuti
<code>query.remote-task.quick-terminate-on-de-failure</code>	Attiva un errore rapido del nodo se il coordinatore non riesce a raggiungere il nodo o non riesce a connettersi al worker in esecuzione su quel nodo. Il valore di <code>query.remote-task.terminate-on-connect-exception</code> determina se il coordinatore deve raggiungere il nodo o connettersi al worker. Il nodo fallisce la query e Amazon EMR lo rimuove dall'elenco dei worker disponibili. In tal caso, non potrai utilizzare il nodo per pianificare nuove query.	true

Impostazione	Descrizione	Default
	Quando imposti questo valore su <code>false</code> , Presto torna al suo comportamento precedente in cui il coordinatore Presto prova nuovamente a raggiungere il nodo (per <code>query.remote-task.max-backoff-duration</code>) prima di contrassegnare il nodo come non disponibile e fallisce la query in corso sul nodo.	
<code>query.remote-task.terminate-on-connection-exception</code>	Specifica se Amazon EMR deve creare un nodo se l'host è raggiungibile ma il coordinatore non riesce a connettersi al processo worker dell'host. Quando imposti questo valore su <code>true</code> , se l'host non è raggiungibile attivi un errore di interrogazione rapida.	<code>false</code>

Esecuzione con tolleranza ai guasti in Trino

L'esecuzione a tolleranza d'errore è un meccanismo di Trino che un cluster può utilizzare per mitigare gli errori delle query. A tale scopo, ritenta le query o le attività dei componenti in caso di errore.

Quando l'esecuzione a tolleranza d'errore è abilitata, i dati intermedi di scambio vengono sottoposti a spool e possono essere riutilizzati da un altro lavoratore in caso di interruzione del lavoratore o di altri errori durante l'esecuzione della query.

Per ulteriori informazioni sull'esecuzione con tolleranza ai guasti in Trino, consulta l'articolo [Project Tardigrade delivers ETL at Trino speeds to early users](#) (Project Tardigrade fornisce ETL alla velocità di Trino ai primi utenti) sul blog di Trino.

Configurazione

L'esecuzione a tolleranza d'errore è disabilitata per impostazione predefinita. Per abilitare la caratteristica, imposta la proprietà di configurazione `retry-policy` nella classificazione `trino-config` su `QUERY` o `TASK` in base alla policy di ripetizione desiderata, come segue.

```
{
  "classification":
    "trino-config",
  "properties":
    {
      "retry-policy":
        "QUERY"
    }
}
```

Una policy di ripetizione **QUERY** indica a Trino di ritentare automaticamente una query in caso di errore su un nodo worker. Ti consigliamo di utilizzare una policy di ripetizione `QUERY` quando la maggior parte del carico di lavoro per il cluster Trino comprende molte query di piccole dimensioni.

Una policy di ripetizione **TASK** indica a Trino di ritentare le singole attività di query in caso di errore. Ti consigliamo di utilizzare questa policy quando Trino esegue query in batch di grandi dimensioni. Il cluster può riprovare in modo più efficiente le attività più piccole all'interno della query anziché riprovare l'intera query.

Exchange Manager

Un Exchange Manager archivia e gestisce i dati sottoposti a spooling per un'esecuzione a tolleranza di errore. Utilizza l'archiviazione esterna per memorizzare i dati distribuiti oltre la dimensione del buffer in memoria. Puoi configurare un Exchange Manager basato su file system che archivia i dati sottoposti a spooling in una posizione specifica, ad esempio Amazon S3, sistemi compatibili con Amazon S3 o HDFS.

Amazon EMR versione 6.9.0 e successive include la classificazione `trino-exchange-manager` per configurare Exchange Manager. Queste versioni supportano anche HDFS per lo spooling.

Configurazione di Exchange Manager

Utilizza la classificazione di configurazione `trino-exchange-manager` per impostare un Exchange Manager. Questa classificazione crea internamente un file di configurazione `etc/exchange-`

`manager.properties` sul coordinatore e su tutti i nodi worker e imposta inoltre la proprietà di configurazione `exchange-manager.name` su `filesystem`.

Per impostazione predefinita, Amazon EMR rilascio 6.9.0 e successivi utilizzano HDFS come gestore di scambio. HDFS è disponibile nei cluster EC2 di Amazon EMR e per impostazione predefinita lo spooling avviene nella directory `trino-exchange/`. Per utilizzare le impostazioni predefinite, imposta la seguente configurazione:

```
{"Classification":
  "trino-exchange-manager"
}
```

Se desideri fornire una posizione personalizzata, imposta le seguenti proprietà nella classificazione di `trino-exchange-manager`:

- Imposta `exchange.use-local-hdfs` su `true`.
- Imposta `exchange.base-directories` sulla posizione della directory personalizzata in HDFS, ad esempio `exchange.base-directories=/exchange`. Se la directory personalizzata non è ancora presente in HDFS, Amazon EMR la creerà.

Configurazioni di Exchange Manager per HDFS

In base ai risultati dei test interni, ti consigliamo di utilizzare lo spooling da HDFS locale per migliorare le prestazioni delle query rispetto ad altri file system basati su cloud. È possibile impostare le seguenti configurazioni per Exchange Manager con HDFS.

Configurazione	Descrizione	Impostazioni predefinite
<code>exchange.hdfs.block-size</code>	Dimensione del blocco per l'archiviazione HDFS	4 MB
<code>hdfs.config.resources</code>	Elenco dei percorsi dei file per configurare HDFS	Se <code>exchange.use-local-hdfs</code> è <code>true</code> , utilizza i percorsi per i file <code>core-site.xml</code> , <code>hdfs-site.xml</code> ; in caso contrario, <code>null</code>

Per ulteriori proprietà di configurazione dell'esecuzione a tolleranza d'errore e per informazioni sulla configurazione di Amazon S3 o di altri sistemi compatibili con Amazon S3 per lo spooling, consulta la pagina [Fault-tolerant execution](#) (Esecuzione a tolleranza d'errore) della documentazione di Trino.

Considerazioni e limitazioni

- Se abiliti l'esecuzione con tolleranza ai guasti, vengono disattivate le operazioni `write` per i connettori che non supportano `write` quando `retry-policy` è impostato. A partire dal rilascio 6.9.0 di Amazon EMR, i connettori Delta Lake, Hive e Iceberg supportano le operazioni `write` con `retry-policy`.
- Se utilizzi Exchange Manager ed esegui operazioni di I/O costose, le prestazioni delle query potrebbero risultare compromesse durante lo spooling dei dati intermedi all'archiviazione esterna da parte di Exchange Manager.

Utilizzo della scalabilità automatica Presto con disattivazione con tolleranza

Le versioni 5.30.0 e successive di Amazon EMR includono una funzionalità che puoi utilizzare per impostare un periodo di tolleranza per determinate operazioni di scalabilità. Il periodo di tolleranza consente alle attività Presto di continuare a essere eseguite prima della terminazione del nodo a causa di un'operazione di ridimensionamento in riduzione o di una richiesta di policy di scalabilità automatica. Per ulteriori informazioni sulle regole di scalabilità, consulta [Comprensione delle regole di scalabilità automatica](#) nella Guida alla gestione di Amazon EMR. La scalabilità automatica Presto con disattivazione con tolleranza impedisce la pianificazione di nuove attività su un nodo in fase di disattivazione, consentendo allo stesso tempo alle attività già in esecuzione di essere completate prima che venga raggiunto il timeout di arresto. Le query in esecuzione verranno completate prima che il nodo venga disattivato. La scalabilità automatica non è supportata nei parchi istanze.

È possibile controllare quanto tempo occorre per il completamento delle attività Presto in seguito alla ricezione di una richiesta di arresto con scalabilità automatica. Per impostazione predefinita, il timeout di arresto per Amazon EMR è di 0 minuti, il che significa che Amazon EMR termina immediatamente il nodo e tutte le attività di Presto in esecuzione, se necessario per una richiesta di riduzione. Per impostare un timeout più lungo per le attività Presto on Amazon EMR e consentire il completamento delle query in esecuzione prima di eseguire il dimensionamento di un cluster, utilizza la classificazione di configurazione `presto-config` per impostare il parametro `graceful-`

`shutdown-timeout` su un valore in secondi o minuti maggiore di zero. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Ad esempio, aumentando il valore `graceful-shutdown-timeout` a `"30m"` si specifica un periodo di timeout di 30 minuti. Al termine del periodo del timeout di arresto, il nodo contrassegnato per la disattivazione viene terminato forzatamente se è in attesa del completamento delle attività di query e la query ha esito negativo. Se le attività di query sono terminate in cinque minuti, il nodo contrassegnato per la disattivazione termina dopo cinque minuti, a condizione che le altre applicazioni YARN abbiano completato l'esecuzione.

Example Esempio di configurazione della scalabilità automatica di Presto con disattivazione con tolleranza

Sostituire il valore `graceful-shutdown-timeout` con il numero di minuti appropriato per la configurazione. Non è previsto un valore massimo. L'esempio seguente imposta un valore di timeout di 1800 secondi (30 minuti).

```
[
  {
    "classification": "presto-config",
    "properties": {
      "graceful-shutdown-timeout": "1800s"
    }
  }
]
```

Limitazioni

La disattivazione con tolleranza di PrestoDB non funziona sui cluster EMR in cui la connettività HTTP è disabilitata, ad esempio quando `http-server.http.enabled` è impostato su `false`. Trino non supporta la disattivazione con tolleranza, indipendentemente dall'impostazione di `http-server.http.enabled`.

Considerazioni riguardanti Presto on Amazon EMR

Considera le seguenti limitazioni quando esegui [Presto](#) su Amazon EMR.

Eseguibile della riga di comando Presto

In Amazon EMR, PrestoDB e Trino utilizzano entrambi lo stesso eseguibile della riga di comando, `presto-cli`, come nell'esempio seguente.

```
presto-cli --catalog hive
```

Proprietà di implementazione Presto non configurabili

La versione di Amazon EMR utilizzata determina le configurazioni di implementazione di Presto disponibili. Per ulteriori informazioni su queste proprietà di configurazione, consulta [Implementazione di Presto](#) nella documentazione Presto. Nella seguente tabella sono riportate le diverse opzioni di configurazione per i file `properties` Presto.

File	Configurabile
<code>log.properties</code>	<p>PrestoDB: configurabile in Amazon EMR versione 4.0.0 e successive. Utilizza la classificazione della configurazione <code>presto-log</code> .</p> <p>Trino (PrestoSQL): configurabile in Amazon EMR versione 6.1.0 e successive. Utilizza la classificazione della configurazione <code>prestoql-log</code> o <code>trino-log</code> .</p>
<code>config.properties</code>	<p>PrestoDB: configurabile in Amazon EMR versione 4.0.0 e successive. Utilizza la classificazione della configurazione <code>presto-config</code> .</p> <p>Trino (PrestoSQL): configurabile in Amazon EMR versione 6.1.0 e successive. Utilizza la classificazione della configurazione <code>prestoql-config</code> o <code>trino-config</code> .</p>
<code>hive.properties</code>	<p>PrestoDB: configurabile in Amazon EMR versione 4.1.0 e successive. Utilizza la classificazione della configurazione <code>presto-connector-hive</code> .</p>

File	Configurabile
	Trino (PrestoSQL): configurabile in Amazon EMR versione 6.1.0 e successive. Utilizza la classificazione della configurazione <code>prestoql-connector-hive</code> o <code>trino-connector-hive</code> .
<code>node.properties</code>	PrestoDB: configurabile in Amazon EMR versione 5.6.0 e successive. Utilizza la classificazione della configurazione <code>presto-node</code> . Trino (PrestoSQL): configurabile in Amazon EMR versione 6.1.0 e successive. Utilizza la classificazione della configurazione <code>prestoql-node</code> o <code>trino-node</code> .
<code>jvm.config</code>	Non configurabile

Installazione di PrestoDB e Trino

Il nome dell'applicazione Presto continua ad essere utilizzato per installare PrestoDB sui cluster. Per installare Trino sui cluster, usa il nome dell'applicazione Trino (o PrestoSQL nelle versioni precedenti di Amazon EMR).

Puoi installare PrestoDB o Trino, ma non puoi installare entrambi in un singolo cluster. Se specifichi sia PrestoDB che Trino quando provi a creare un cluster, si verifica un errore di convalida e la richiesta di creazione del cluster ha esito negativo.

Configurazione EMRFS e PrestoS3FileSystem

Con Amazon EMR versione 5.12.0 e successive, PrestoDB può utilizzare EMRFS. Questa è la configurazione predefinita. EMRFS è anche il file system predefinito per Trino (PrestoSQL) in Amazon EMR versione 6.1.0 e successive. Per ulteriori informazioni, consulta [File System EMR \(EMRFS\)](#) nella Guida alla gestione di Amazon EMR. Con le versioni precedenti di Amazon EMR, PrestoS3FileSystem è l'unica opzione di configurazione.

Puoi utilizzare una configurazione di sicurezza per impostare la crittografia per i dati EMRFS in Amazon S3. Puoi anche utilizzare ruoli IAM per richieste EMRFS ad Amazon S3. Per ulteriori

informazioni, consulta [Comprensione delle opzioni di crittografia](#) e [Configurazione dei ruoli IAM per richieste EMRFS ad Amazon S3](#) nella Guida alla gestione di Amazon EMR.

Note

Se interroghi i dati sottostanti in Amazon S3 con Amazon EMR versione 5.12.0, possono verificarsi errori di Presto. Questo avviene perché Presto non riesce a raccogliere i valori di classificazione della configurazione da `emrfs-site.xml`. Come soluzione alternativa, crea una sottodirectory `emrfs` in `usr/lib/presto/plugin/hive-hadoop2/` e crea un collegamento simbolico in `usr/lib/presto/plugin/hive-hadoop2/emrfs` sul file `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` esistente. Quindi riavvia il processo `presto-server` (`sudo presto-server stop` seguito da `sudo presto-server start`).

Puoi ignorare l'impostazione predefinita EMRFS e utilizzare invece `PrestoS3FileSystem`. A questo scopo, utilizza la classificazione della configurazione `presto-connector-hive` per impostare `hive.s3-file-system-type` su `PRESTO` come mostrato nel seguente esempio. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.s3-file-system-type": "PRESTO"
    }
  }
]
```

Se utilizzi `PrestoS3FileSystem`, utilizza la classificazione di configurazione `presto-connector-hive` o `trino-connector-hive` per Trino per configurare le proprietà di `PrestoS3FileSystem`. Per ulteriori informazioni sulle proprietà disponibili, consulta [Configurazione di Amazon S3](#) nella sezione Connettore Hive della documentazione Presto. Queste impostazioni non si applicano a EMRFS.

Impostazione predefinita per rappresentazione utente finale

Per impostazione predefinita, Amazon EMR versione 5.12.0 e successive consente la rappresentazione dell'utente finale per l'accesso a HDFS. Per ulteriori informazioni, consulta

la sezione relativa alla [End user impersonation \(Rappresentazione dell'utente finale\)](#) nella documentazione di Presto. Puoi modificare questa impostazione utilizzando la classificazione di configurazione presto-config impostando la proprietà `hive.hdfs.impersonation.enabled` su `false`.

Porta predefinita per interfaccia Web Presto

Per impostazione predefinita, Amazon EMR configura l'interfaccia Web Presto sul sistema di coordinamento Presto per utilizzare la porta 8889 (per PrestoDB e Trino). Per modificare la porta, utilizza la classificazione di configurazione presto-config per impostare la proprietà `http-server.http.port`. Per ulteriori informazioni, consulta [Config properties \(Proprietà di configurazione\)](#) nella sezione Deploying Presto (Implementazione di Presto) della documentazione di Presto.

Problema con l'esecuzione di bucket Hive in alcuni rilasci

Presto versione 152.3 ha un problema con l'esecuzione del bucket Hive che riduce notevolmente le prestazioni delle query Presto in alcune circostanze. Questa versione di Presto è inclusa nelle versioni di Amazon EMR 5.0.3, 5.1.0 e 5.2.0. Per ridurre il verificarsi di questo problema, utilizza la classificazione della configurazione presto-connector-hive per impostare la proprietà `hive.bucket-execution` su `false` come mostrato nel seguente esempio.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.bucket-execution": "false"
    }
  }
]
```

Cronologia delle versioni di Presto

La seguente tabella indica la versione di Presto inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni relative alla versione Presto

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
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
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-

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
		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
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

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.10.1	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.10.0	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.9.1	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.9.0	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.8.1	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.8.0	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.7.0	0.272	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.1	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.36.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.6.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.35.0	0.266	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.5.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.4.0	0.254.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.3.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.3.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.2.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.2.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.1.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.1.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.0.1	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-6.0.0	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.34.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.33.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.33.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.32.1	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.32.0	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.31.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.31.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.30.2	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.30.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.29.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.28.1	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.27.1	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.27.0	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.26.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.25.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.24.1	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.0	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.23.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.23.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.22.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.2	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.21.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.20.1	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.20.0	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.19.1	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.19.0	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.18.1	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.0	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.17.2	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.1	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.17.0	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.16.1	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.16.0	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.15.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.15.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.2	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.14.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.13.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.12.3	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.2	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.12.1	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.0	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.11.4	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.3	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.11.2	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.11.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.10.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.10.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.9.1	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.9.0	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.8.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.8.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.7.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.7.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.6.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.4	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.5.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.4.1	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.4.0	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.3.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.0	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.2.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
emr-5.1.1	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.1.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.3	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Etichetta di rilascio di Amazon EMR	Versione di Presto	Componenti installati con Presto
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

La seguente tabella indica la versione di Trino (Presto SQL) inclusa in ogni versione del rilascio di Amazon EMR e i componenti installati con l'applicazione. PrestoSQL ha cambiato il suo nome in Trino a partire dalla versione 351.

Informazioni sulla versione di Trino (PrestoSQL)

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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,

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.12.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
emr-6.9.0	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.8.1	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
emr-6.8.0	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.7.0	378	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Etichetta di rilascio di Amazon EMR	Versione di Trino (PrestoSQL)	Componenti installati con Trino (PrestoSQL)
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

Note di rilascio di Trino (PrestoSQL) per versione

- [Amazon EMR 6.9.0 - Note sul rilascio per Trino \(PrestoSQL\)](#)

Amazon EMR 6.9.0 - Note sul rilascio per Trino (PrestoSQL)

Amazon EMR 6.9.0 - Nuove caratteristiche di Trino (PrestoSQL)

- Per supportare le query di lunga durata, Trino ora include un meccanismo di esecuzione con tolleranza ai guasti. L'esecuzione con tolleranza ai guasti mitiga gli errori delle query ritentando le query non riuscite o le attività dei loro componenti. Per ulteriori informazioni, consulta [Esecuzione con tolleranza ai guasti in Trino](#).

Amazon EMR 6.9.0 - Modifiche a Trino (PrestoDB)

Amazon EMR 6.9.0 - Modifiche a PrestoDB

Tipo	Descrizione
Upgrade	Aggiornamento di Presto a 0.276

Tipo	Descrizione
Upgrade	Supporto per Hadoop 3.3.3
Upgrade	Aggiornamento di Hudi a 0.12.1
Caratteristica	Integrazione di Amazon EMR e Presto con AWS Lake Formation per carichi di lavoro interattivi utilizzando l'API di GCSC.
Caratteristica	Sono state aggiunte configurazioni relative a Kerberos all'interno delle configurazioni di sicurezza di PrestoDB per abilitare Keberos.
Correzioni di bug	È stata annullata la richiesta di pull OSS #18115, che era stata aggiunta per ridurre il numero di copie <code>hdfsConfiguration</code> . Ciò causava una copia errata della configurazione HDFS con l'utilizzo di tabelle EMRFS o Hudi.

Amazon EMR 6.9.0 - Modifiche a Trino

Tipo	Descrizione
Upgrade	Aggiornamento di Trino a 398
Upgrade	Supporto per Hadoop 3.3.3
Caratteristica	Supporto Tardigrade: è stato aggiunto il supporto per lo spooling di scambio su HDFS e Amazon S3. Per ulteriori informazioni, consulta Esecuzione con tolleranza ai guasti in Trino
Correzioni di bug	Quando utilizzi Trino Iceberg e Glue Catalog è abilitato, evita di aggiungere l'URI del metastore in <code>iceberg.properties</code> .

Amazon EMR 6.9.0 - Problemi noti di Trino (PrestoSQL)

- Per Amazon EMR rilascio 6.9.0, Trino non funziona su cluster abilitati per Apache Ranger. Se hai la necessità di utilizzare Trino con Ranger, contatta il [AWS Support](#).

Apache Spark

[Apache Spark](#) è un framework di elaborazione distribuito e un modello di programmazione che consente di eseguire attività come machine learning, elaborazione di flussi o analisi di grafici con i cluster Amazon EMR. Come Apache Hadoop, Spark è un sistema di elaborazione distribuito open source utilizzato in genere per carichi di lavoro di Big Data. Presenta tuttavia importanti differenze rispetto a Hadoop MapReduce, in quanto dispone di un motore di esecuzione di grafi aciclici orientati (DAG) ottimizzato e memorizza attivamente i dati nella cache. Queste caratteristiche possono migliorare le prestazioni, in particolare per determinati algoritmi e query interattive.

Spark supporta applicazioni scritte in Scala, Python e Java a livello nativo. Include inoltre diverse librerie strettamente integrate per SQL ([Spark SQL](#)), machine learning ([MLlib](#)), elaborazione di flussi ([Spark Streaming](#)) e analisi di grafici ([GraphX](#)). Questi strumenti semplificano l'utilizzo del framework Spark per un'ampia gamma di casi d'uso.

Puoi installare Spark su un cluster Amazon EMR insieme ad altre applicazioni Hadoop. Questa soluzione consente inoltre l'utilizzo del file system EMR (EMRFS) per accedere direttamente ai dati in Amazon S3. Spark integra inoltre Hive di modo che sia possibile utilizzare un oggetto HiveContext per eseguire script Hive utilizzando Spark. Un contesto Hive è incluso nella shell di Spark come `sqlContext`.

Per un tutorial su come configurare un cluster EMR con Spark e analizzare un set di dati di esempio, consulta [Tutorial: Getting started with Amazon EMR](#) (Tutorial: nozioni di base di Amazon EMR) sul blog AWS News.

Important


Apache Spark versione 2.3.1, disponibile a partire da Amazon EMR rilascio 5.16.0, risolve [CVE-2018-8024](#) e [CVE-2018-1334](#). Si consiglia di eseguire la migrazione delle versioni precedenti di Spark a Spark 2.3.1 o versioni successive.

La seguente tabella indica la versione di Spark inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Spark.

Per la versione dei componenti installati con Spark in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Spark per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.14.0	Spark 3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

 Note

Rilascio 6.8.0 di Amazon EMR con Spark per Apache 3.3.0. Questa versione di Spark utilizza Apache Log4j 2 e il file `log4j2.properties` per configurare Log4j nei processi Spark. Se utilizzi Spark nel cluster o crei cluster EMR con parametri di configurazione personalizzati e desideri eseguire l'aggiornamento alla versione 6.8.0 di Amazon EMR, devi migrare alla nuova classificazione della configurazione e formato chiave `spark-log4j2` per Apache Log4j 2. Per ulteriori informazioni, consulta [Migrazione da Apache Log4j 1.x a Log4j 2.x](#).

La seguente tabella indica la versione di Spark inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Spark.

Per la versione dei componenti installati con Spark in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Spark per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.36.1	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-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

Argomenti

- [Creazione di un cluster con Spark](#)
- [Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.x](#)
- [Utilizzare AWS Glue Data Catalog come metastore Spark SQL](#)
- [Configurazione di Spark](#)
- [Ottimizzazione delle prestazioni di Spark](#)
- [Caching dei frammenti dei risultati Spark](#)
- [Utilizzare l'Nvidia Spark-RAPIDS Accelerator per Spark](#)
- [Accesso alla shell Spark](#)
- [Utilizzare Amazon SageMaker Spark per il machine learning](#)
- [Scrittura di un'applicazione Spark](#)
- [Migliorare le prestazioni di Spark con Amazon S3](#)
- [Aggiungi una fase Spark](#)

- [Visualizzazione della cronologia delle applicazioni Spark](#)
- [Accedi alle interfacce utente Web di Spark](#)
- [Utilizzo dell'integrazione di Amazon Redshift per Apache Spark su Amazon EMR](#)
- [Cronologia dei rilasci di Spark](#)

Creazione di un cluster con Spark

La procedura seguente consente di creare un cluster con [Spark](#) installato utilizzando Quick Options (Opzioni rapide) nella console di EMR.

In alternativa, puoi utilizzare Advanced Options (Opzioni avanzate) per personalizzare ulteriormente la configurazione del cluster o per inviare fasi al fine di installare applicazioni a livello di codice e quindi eseguire le applicazioni personalizzate. Con una di queste opzioni di creazione del cluster, puoi scegliere di utilizzare AWS Glue come metastore Spark SQL. Per ulteriori informazioni, consulta [Utilizzare AWS Glue Data Catalog come metastore Spark SQL](#).

Per avviare un cluster con Spark installato

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. Scegli Create cluster (Crea cluster) per utilizzare Quick Options (Opzioni rapide).
3. Immetti un Cluster name (Nome cluster).
4. Per Software Configuration (Configurazione del software), scegli un'opzione di Release (Rilascio).
5. Per Applications (Applicazioni), scegli il bundle di applicazioni Spark.
6. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Note

Per configurare Spark durante la creazione del cluster, consulta [Configurazione di Spark](#).

Per avviare un cluster con Spark installato mediante il kit AWS CLI

- Creare il cluster con il seguente comando:

```
aws emr create-cluster --name "Spark cluster" --release-label emr-5.36.1 --  
applications Name=Spark \
```

```
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Avvio di un cluster con Spark installato mediante il kit SDK per Java

Specificare Spark come applicazione con `SupportedProductConfig` utilizzato in `RunJobFlowRequest`.

- L'esempio seguente mostra come creare un cluster con Spark utilizzando Java.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

    public static void main(String[] args) {
        AWSCredentials credentials_profile = null;
        try {
            credentials_profile = new
ProfileCredentialsProvider("default").getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                "Cannot load credentials from .aws/credentials file. " +
                "Make sure that the credentials file exists and the profile
name is specified within it.",
                e);
        }
    }
}
```

```
AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
    .withRegion(Regions.US_WEST_1)
    .build();

// create a step to enable debugging in the AWS Management Console
StepFactory stepFactory = new StepFactory();
StepConfig 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());
}
}
```

Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.x

Con Amazon EMR 6.0.0, le applicazioni Spark possono utilizzare i container Docker per definire le dipendenze delle librerie anziché installare le dipendenze sulle singole istanze Amazon EC2

del cluster. Per eseguire Spark con Docker, è necessario innanzitutto configurare il Registro di sistema Docker e definire parametri aggiuntivi quando si invia un'applicazione Spark. Per ulteriori informazioni, consulta [Configurazione dell'integrazione di Docker](#).

Quando l'applicazione viene inviata, YARN richiama Docker per estrarre l'immagine Docker specificata ed eseguire l'applicazione Spark all'interno di un container Docker. Ciò consente di definire e isolare facilmente le dipendenze. Riduce il tempo per il processo di bootstrap o la preparazione delle istanze nel cluster Amazon EMR con le librerie necessarie per l'esecuzione del processo.

Considerazioni sull'esecuzione di Spark con Docker

Quando si esegue Spark con Docker, assicurarsi che siano soddisfatti i seguenti prerequisiti:

- Il pacchetto `docker` e l'interfaccia della CLI sono installati solo sui nodi core e task.
- In alternativa, su Amazon EMR rilascio 6.1.0 e successivi è possibile installare Docker su un nodo primario utilizzando i seguenti comandi.

```
sudo yum install -y docker
sudo systemctl start docker
```

- Il comando `spark-submit` deve sempre essere eseguito da un'istanza primaria nel cluster Amazon EMR.
- I registri Docker utilizzati per risolvere le immagini Docker devono essere definiti utilizzando l'API di classificazione con la chiave di classificazione `container-executor` per definire parametri aggiuntivi all'avvio del cluster:
 - `docker.trusted.registries`
 - `docker.privileged-containers.registries`
- Per eseguire un'applicazione Spark in un container Docker, sono necessarie le seguenti opzioni di configurazione:
 - `YARN_CONTAINER_RUNTIME_TYPE=docker`
 - `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- Quando si utilizza Amazon ECR per recuperare le immagini Docker, è necessario configurare il cluster per autenticarsi. A tale scopo, è necessario utilizzare la seguente opzione di configurazione:
 - `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON`

- In EMR 6.1.0 e versioni successive non è necessario utilizzare il comando elencato `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` quando è abilitata la funzionalità di autenticazione automatica ECR.
- Qualsiasi immagine Docker utilizzata con Spark deve avere Java installato nell'immagine Docker.

Per ulteriori informazioni sui prerequisiti, consulta [Configurazione dell'integrazione di Docker](#).

Creazione di un'immagine Docker

Le immagini Docker vengono create utilizzando un Dockerfile, che definisce i pacchetti e la configurazione da includere nell'immagine. I seguenti due Dockerfiles di esempio utilizzano PySpark e SparkR.

Dockerfile di PySpark

Le immagini Docker create da questo Dockerfile includono Python 3 e il pacchetto NumPy Python. Questo Dockerfile utilizza Amazon Linux 2 e Amazon Corretto JDK 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"
```

Dockerfile SparkR

Le immagini Docker create da questo Dockerfile includono R e il pacchetto RandomForest CRAN. Questo Dockerfile include Amazon Linux 2 e Amazon Corretto JDK 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')"
```

Per ulteriori informazioni sulla sintassi Dockerfile, vedere la [documentazione di riferimento Dockerfile](#).

Utilizzo delle immagini Docker da Amazon ECR

Amazon Elastic Container Registry (Amazon ECR) è un registro completamente gestito del container Docker che permette agli sviluppatori di archiviare, gestire e distribuire in modo semplice le immagini dei container Docker. Durante l'utilizzo di Amazon ECR, il cluster deve essere configurato in modo da considerare attendibile l'istanza di ECR ed è necessario configurare l'autenticazione affinché il cluster possa utilizzare le immagini Docker da Amazon ECR. Per ulteriori informazioni, consulta [Configurazione di YARN per accedere ad Amazon ECR](#).

Per assicurarsi che gli host EMR possano accedere alle immagini archiviate in Amazon ECR, il cluster deve disporre delle autorizzazioni dalla policy `AmazonEC2ContainerRegistryReadOnly` associata al profilo dell'istanza. Per maggiori informazioni, consulta [Policy AmazonEC2ContainerRegistryReadOnly](#).

In questo esempio, il cluster deve essere creato con la seguente configurazione aggiuntiva per garantire che il registro Amazon ECR sia attendibile. Sostituisci l'endpoint `123456789123.dkr.ecr.us-east-1.amazonaws.com` con l'endpoint Amazon ECR.

```
[
  {
    "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": {}  
  }  
]
```

Utilizzo di PySpark con Amazon ECR

Nell'esempio seguente viene utilizzato il file Dockerfile PySpark, che verrà taggato e caricato su Amazon ECR. Dopo il caricamento del Dockerfile, è possibile eseguire il processo PySpark e fare riferimento all'immagine Docker da Amazon ECR.

Dopo aver avviato il cluster, utilizzare SSH per connettersi a un nodo principale ed eseguire i comandi seguenti per creare l'immagine Docker locale dal Dockerfile PySpark di esempio.

Innanzitutto, crea una directory e un Dockerfile.

```
mkdir pyspark  
vi pyspark/Dockerfile
```

Incollare il contenuto del Dockerfile PySpark ed eseguire i seguenti comandi per creare un'immagine Docker.

```
sudo docker build -t local/pyspark-example pyspark/
```

Creare il repository `emr-docker-examples` ECR per gli esempi.

```
aws ecr create-repository --repository-name emr-docker-examples
```

Etichetta e carica l'immagine creata localmente in ECR, sostituendo `123456789123.dkr.ecr.us-east-1.amazonaws.com` con l'endpoint ECR.

```
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
```

Utilizza SSH per connetterti al nodo primario e preparare uno script Python con il nome del file `main.py`. Incolla il seguente contenuto nel file `main.py` e salvalo.

```
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)
```

Per inviare il processo su EMR 6.0.0, fare riferimento al nome dell'immagine Docker. Definire i parametri di configurazione aggiuntivi per assicurarsi che l'esecuzione del processo utilizzi Docker come runtime. Durante l'utilizzo di Amazon ECR, `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` deve fare riferimento al file `config.json` contenente le credenziali utilizzate per l'autenticazione ad Amazon ECR.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=$DOCKER_CLIENT_CONFIG \
--num-executors 2 \
main.py -v
```

Per inviare il processo su EMR 6.1.0 e versioni successive, fai riferimento al nome dell'immagine Docker. Quando è attivata l'autenticazione automatica ECR, esegui il comando riportato di seguito.

```
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
```

Al termine del processo, prendere nota dell'ID applicazione YARN e utilizzare il comando seguente per ottenere l'output del processo PySpark.

```
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]]
```

Utilizzo di SparkR con Amazon ECR

Nell'esempio seguente viene utilizzato il Dockerfile SparkR, che verrà taggato e caricato in ECR. Una volta caricato il Dockerfile, è possibile eseguire il processo SparkR e fare riferimento all'immagine Docker da Amazon ECR.

Dopo aver avviato il cluster, utilizzare SSH per connettersi a un nodo principale ed eseguire i comandi seguenti per creare l'immagine Docker locale dall'esempio Dockerfile SparkR.

Innanzitutto, creare una directory e il Dockerfile.

```
mkdir sparkr  
vi sparkr/Dockerfile
```

Incollare il contenuto del Dockerfile SparkR ed eseguire i seguenti comandi per creare un'immagine Docker.

```
sudo docker build -t local/sparkr-example sparkr/
```

Etichetta e carica l'immagine creata localmente in Amazon ECR, sostituendo `123456789123.dkr.ecr.us-east-1.amazonaws.com` con l'endpoint Amazon ECR.

```
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
```

Utilizza SSH per connetterti al nodo primario e preparare uno script R con il nome `sparkR.R`. Incollare i contenuti seguenti nel file `sparkR.R`.

```
library(SparkR)
sparkR.session(appName = "R with Spark example", sparkConfig =
  list(spark.some.config.option = "some-value"))

sqlContext <- sparkRSQL.init(spark.sparkContext)
library(randomForest)
# check release notes of randomForest
rfNews()

sparkR.session.stop()
```

Per inviare il processo su EMR 6.0.0, fai riferimento al nome dell'immagine Docker. Definire i parametri di configurazione aggiuntivi per assicurarsi che l'esecuzione del processo utilizzi Docker come runtime. Durante l'utilizzo di Amazon ECR, `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` deve fare riferimento al file `config.json` contenente le credenziali utilizzate per l'autenticazione a 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
```

Per inviare il processo su EMR 6.1.0 e versioni successive, fai riferimento al nome dell'immagine Docker. Quando è attivata l'autenticazione automatica ECR, esegui il comando riportato di seguito.

```
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
```

Una volta completato il processo, prendere nota dell'ID applicazione YARN e utilizzare il comando seguente per ottenere l'output del processo SparkR. Questo esempio include test per assicurarsi che la libreria RandomForest, la versione installata e le note sulla versione siano disponibili.

```
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:
```

Utilizzare AWS Glue Data Catalog come metastore Spark SQL

Con Amazon EMR rilascio 5.8.0 o successivi, è possibile configurare Spark SQL per utilizzare AWS Glue Data Catalog come metastore. Si consiglia questa configurazione quando si richiede un metastore persistente o un metastore condiviso da diversi cluster, servizi e applicazioni o account AWS.

AWS Glue è un servizio di estrazione, trasformazione e caricamento (ETL) completamente gestito che consente di catalogare, filtrare e migliorare la qualità dei dati in modo semplice e conveniente, oltre a trasferirli in modo sicuro tra datastore. AWS Glue Data Catalog offre un repository di metadati unificato per varie origini dati e formati di dati, integrandosi con Amazon EMR così come con Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena e qualsiasi applicazione compatibile con il metastore Apache Hive. AWS I crawler di Glue possono dedurre automaticamente lo schema dai dati di origine in Amazon S3 e archiviare i metadati associati nel Data Catalog. Per ulteriori informazioni sul Data Catalog, consulta la sezione relativa al [Popolamento di AWS Glue Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

AWS Glue è soggetto a costi separati. Viene applicata una tariffa mensile per l'archiviazione e l'accesso ai metadati nel Data Catalog, una tariffa oraria fatturata al minuto per i processi ETL di AWS Glue e per il runtime del crawler e, infine, una tariffa oraria fatturata al minuto per ciascun endpoint di sviluppo di cui è stato effettuato il provisioning. Il Data Catalog consente di archiviare fino a un milione di oggetti senza costi aggiuntivi. Se intendi archiviare più di un milione di oggetti, ti verrà addebitato 1 USD per ogni 100.000 oggetti oltre il milione. Un oggetto nel Data Catalog può essere una tabella, una partizione o un database. Per ulteriori informazioni, consulta [Prezzi di Glue](#).

Important

Se hai creato tabelle utilizzando Amazon Athena o Amazon Redshift Spectrum prima del 14 agosto 2017, i database e le tabelle sono archiviati in un catalogo gestito da Athena, che è separato da AWS Glue Data Catalog. Per integrare Amazon EMR con queste tabelle, dovrai eseguire l'aggiornamento a AWS Glue Data Catalog. Per ulteriori informazioni, consulta [Aggiornamento a AWS Glue Data Catalog](#) nella Guida per l'utente di Amazon Athena.

Specifica di AWS Glue Data Catalog come metastore

È possibile specificare AWS Glue Data Catalog come metastore utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Quando si utilizza la CLI o l'API, è possibile utilizzare la classificazione di configurazione per Spark per specificare un Data Catalog. Inoltre, con Amazon EMR 5.16.0 e versioni successive, è possibile utilizzare la classificazione di configurazione per specificare un Data Catalog in un Account AWS diverso. Quando utilizzi la console, puoi specificare il Data Catalog utilizzando Advanced Options (Opzioni avanzate) o Quick Options (Opzioni rapide).

 Note

L'opzione di utilizzare AWS Glue Data Catalog è disponibile anche con Zeppelin in quanto Zeppelin è installato con i componenti di Spark SQL.

New console

Specifica di AWS Glue Data Catalog come metastore Spark con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Application bundle (Bundle di applicazioni), scegli Spark o Custom (Personalizzato). Se personalizzi il cluster, assicurati di selezionare Zeppelin o Spark come una delle tue applicazioni.
4. In Impostazioni del Catalogo dati AWS Glue, seleziona la casella di controllo Utilizza per i metadati delle tabelle Spark.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di AWS Glue Data Catalog come metastore Spark con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Per Release (Versione), scegliere emr-5.8.0 o versioni successive.
4. In Release (Versione), selezionare Spark o Zeppelin.
5. In AWS Glue Data Catalog settings (Impostazioni Glue Data Catalog), seleziona Use for Spark table metadata (Usa per metadati tabella Spark).
6. Scegliere altre opzioni per il cluster come appropriato, scegliere Next (Avanti), quindi configurare altre opzioni di cluster come appropriato per l'applicazione.

AWS CLI

Specifica di AWS Glue Data Catalog come metastore Spark con la AWS CLI

Per ulteriori informazioni su come specificare una classificazione di configurazione utilizzando l'AWS CLI e l'API EMR, consulta [Configurazione delle applicazioni](#).

- Specificare il valore per `hive.metastore.client.factory.class` utilizzando la classificazione `spark-hive-site` come mostrato nell'esempio seguente:

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Per specificare un Data Catalog in un altro account AWS, aggiungi la proprietà `hive.metastore.glue.catalogid` come illustrato nel seguente esempio. Sostituisci *acct-id* con l'account AWS di Data Catalog.

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```


Autorizzazioni IAM

Il profilo dell'istanza EC2 per un cluster deve disporre di autorizzazioni IAM per le operazioni AWS Glue. Inoltre, se abiliti la crittografia per oggetti AWS Glue Data Catalog, al ruolo deve essere consentito anche crittografare, decrittare e generare la AWS KMS key utilizzata per la crittografia dei dati.

Autorizzazioni per operazioni AWS Glue

Se utilizzi il profilo dell'istanza EC2 di default per Amazon EMR, non è richiesta alcuna operazione. La policy gestita `AmazonElasticMapReduceforEC2Role` collegata al `EMR_EC2_DefaultRole` consente tutte le operazioni AWS Glue necessarie. Tuttavia, se specifichi autorizzazioni e un profilo dell'istanza EC2 personalizzati, devi configurare le operazioni AWS Glue appropriate. Utilizza la policy `AmazonElasticMapReduceforEC2Role` gestita come punto di partenza. Per ulteriori informazioni, consulta [Ruolo di servizio per le istanze EC2 del cluster \(profilo dell'istanza EC2\)](#) nella Guida alla gestione di Amazon EMR.

Autorizzazioni per crittografare e decrittografare AWS Glue Data Catalog

Il tuo profilo dell'istanza ha bisogno dell'autorizzazione per crittografare e decrittare i dati utilizzando la tua chiave. Non è necessario configurare queste autorizzazioni se si applicano entrambe le seguenti istruzioni:

- Per abilitare la crittografia per oggetti AWS Glue Data Catalog utilizzare chiavi gestite per AWS Glue.
- Si utilizza un cluster che è nello stesso Account AWS come il AWS Glue Data Catalog.

In caso contrario, devi aggiungere la istruzione seguente alla policy delle autorizzazioni collegata al profilo dell'istanza EC2.

```
[
  {
    "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"
    }
  ]
}
]

```

Per ulteriori informazioni sulla crittografia di AWS Glue Data Catalog, consulta la sezione [Crittografia del proprio Data Catalog](#) nella Guida per gli sviluppatori di AWS Glue.

Autorizzazioni basate su risorse

Se utilizzi AWS Glue in combinazione con Hive, Spark o Presto in Amazon EMR, AWS Glue supporta policy basate su risorse per controllare l'accesso alle risorse del Data Catalog. Queste risorse includono tabelle, database, connessioni e funzioni definite dall'utente. Per ulteriori informazioni, consulta [Policy sulle risorse AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.

Quando utilizzi policy basate su risorse per limitare l'accesso ad AWS Glue da Amazon EMR, l'entità principale specificata nella policy delle autorizzazioni deve essere il ruolo ARN associato al profilo di istanza EC2 specificato al momento della creazione di un cluster. Ad esempio, per una policy basata sulle risorse collegata a un catalogo, è possibile specificare il ruolo ARN per il ruolo di servizio predefinito per le istanze del cluster EC2, *EMR_EC2_DefaultRole*, come Principal utilizzando il formato mostrato nell'esempio seguente:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

L'*acct-id* può essere diverso dall'ID dell'account AWS Glue. Ciò consente l'accesso da cluster EMR in account diversi. Puoi specificare più entità principali, ognuna da un account diverso.

Considerazioni sull'utilizzo di AWS Glue Data Catalog

Considera i seguenti elementi durante l'utilizzo di AWS Glue Data Catalog come metastore con Spark:

- Disporre di un database predefinito senza l'URI del percorso genera errori durante la creazione di una tabella. Per risolvere il problema, utilizza la clausola LOCATION per specificare un percorso di bucket, ad esempio *s3://EXAMPLE-DOC-BUCKET*, quando utilizzi CREATE TABLE. In alternativa, crea tabelle in un database che non sia quello predefinito.

- La ridenominazione delle tabelle in AWS Glue non è supportata.
- Quando si crea una tabella Hive senza specificare una `LOCATION`, i dati nella tabella vengono archiviati nella posizione specificata dalla proprietà `hive.metastore.warehouse.dir`. Per impostazione predefinita, questa è una posizione in HDFS. Se un altro cluster necessita di accedere alla tabella, non ci riuscirà a meno che non disponga di autorizzazioni adeguate per il cluster che ha creato la tabella. Inoltre, poiché l'archiviazione HDFS è transitoria, se il cluster viene terminato, i dati della tabella vanno persi e la tabella deve essere ricreata. Ti consigliamo di specificare una `LOCATION` in Amazon S3 quando crei una tabella Hive utilizzando AWS Glue. In alternativa, puoi utilizzare la classificazione di configurazione `hive-site` per specificare una posizione in Amazon S3 per `hive.metastore.warehouse.dir`, che si applica a tutte le tabelle Hive. Se una tabella viene creata in una posizione HDFS e il cluster che l'ha creata è ancora in esecuzione, puoi aggiornare la posizione della tabella ad Amazon S3 in AWS Glue. Per ulteriori informazioni, consulta [Utilizzo di tabelle sulla console AWS Glue](#) nella Guida per gli sviluppatori di AWS Glue.
- I valori delle partizioni contenenti virgolette e apostrofi non sono supportati, ad esempio `PARTITION (owner="Doe 's")..`
- Le [statistiche di colonna](#) sono supportate per `emr-5.31.0` e versioni successive.
- L'utilizzo dell'[autorizzazione Hive](#) non è supportato. In alternativa, puoi prendere in considerazione l'utilizzo di [Policy basate sulle risorse AWS Glue](#). Per ulteriori informazioni, consulta [Utilizzo delle policy basate sulle risorse per l'accesso di Amazon EMR a AWS Glue Data Catalog](#).

Configurazione di Spark

Puoi configurare [Spark su Amazon EMR](#) con le classificazioni di configurazione. Per ulteriori informazioni sulle classificazioni di configurazione, consulta [Configurazione delle applicazioni](#).

Le classificazioni di configurazione per Spark su Amazon EMR includono le seguenti:

- **spark**: imposta la proprietà `maximizeResourceAllocation` su `true` (vero) o `false` (falso). Se è `true` (vero), Amazon EMR configura automaticamente le proprietà `spark-defaults` in base alla configurazione hardware del cluster. Per ulteriori informazioni, consulta [Utilizzo di maximizeResourceAllocation](#).
- **spark-defaults**: imposta valori nel file `spark-defaults.conf`. Per ulteriori informazioni, consulta [Spark configuration \(Configurazione di Spark\)](#) nella documentazione di Spark.
- **spark-env**: imposta valori nel file `spark-env.sh`. Per ulteriori informazioni, consulta [Environment variables \(Variabili di ambiente\)](#) nella documentazione di Spark.

- **spark-hive-site**: imposta valori nel file `hive-site.xml` per Spark.
- **spark-log4j**— (Amazon EMR rilascia 6.7.x e versioni precedenti) Imposta i valori nel file `log4j.properties`. Per ulteriori informazioni, consulta il file [log4j.properties.template](#) su Github.
- **spark-log4j2**— (Amazon EMR release 6.8.0 e successive) Imposta i valori nel file `log4j2.properties`. Per ulteriori informazioni, consulta il file [log4j2.properties.template](#) su Github.
- **spark-metrics**: imposta valori nel file `metrics.properties`. Per impostazioni e ulteriori informazioni, consulta il file [metrics.properties.template](#) su Github e la sezione relativa ai [parametri](#) nella documentazione Spark.

Note

Se effettui la migrazione dei carichi di lavoro Spark ad Amazon EMR da un'altra piattaforma, ti consigliamo di testare i carichi di lavoro con [Valori di default di Spark impostati da Amazon EMR](#) prima di aggiungere configurazioni personalizzate. La maggior parte dei clienti vede prestazioni migliorate con le nostre impostazioni di default.


Argomenti


- [Valori di default di Spark impostati da Amazon EMR](#)
- [Configurazione della garbage collection di Spark su Amazon EMR 6.1.0](#)
- [Utilizzo di maximizeResourceAllocation](#)
- [Configurazione del comportamento di disattivazione di nodi](#)
- [Variabile di ambiente ThriftServer Spark](#)
- [Modifica delle impostazioni di default di Spark](#)
- [Migrazione da Apache Log4j 1.x a Log4j 2.x](#)

Valori di default di Spark impostati da Amazon EMR

Nella tabella seguente viene mostrato in che modo Amazon EMR imposta i valori predefiniti in `spark-defaults` che influenzano le applicazioni.

Valori di default di Spark impostati da Amazon EMR

Impostazione	Descrizione	Valore predefinito
<code>spark.executor.memory</code>	La quantità di memoria da utilizzare per processo esecutore. Ad esempio, 1g, 2g.	Questa impostazione è determinata dai tipi di istanza principali e dell'attività presenti nel cluster.
<code>spark.executor.cores</code>	Il numero di core da utilizzare in ogni esecutore.	Questa impostazione è determinata dai tipi di istanza principali e dell'attività presenti nel cluster.
<code>spark.dynamicAllocation.enabled</code>	Se vero, utilizza l'assegnazione di risorse dinamiche, che dimensiona il numero di esecutori registrati con un'applicazione in base al carico di lavoro.	<p>true (con Amazon EMR 4.4.0 e versioni successive)</p> <div data-bbox="1068 871 1510 1234" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; background-color: #e6f2ff;"> <p> Note</p> <p>Il servizio di mescolamento Spark è configurato automaticamente da Amazon EMR.</p> </div>
<code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code>	Se vero, è abilitato il pushdown avanzato dei predicati di partizione nel metastore Hive.	true
<code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code>	Esegue il pushdown dei filtri <code>startsWith</code> , <code>contains</code> e <code>endsWith</code> nel metastore di Hive.	true

Impostazione	Descrizione	Valore predefinito
	<div data-bbox="591 212 1031 955" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px;"> <p> Note</p> <p>Glue non supporta il pushdown del predicato per <code>startsWith</code>, <code>contains</code> o <code>endsWith</code>. Se utilizzi il metastore Glue e riscontri errori dovuti al pushdown dei predicati per queste funzioni, imposta questa configurazione su <code>false</code>.</p> </div>	

Configurazione della garbage collection di Spark su Amazon EMR 6.1.0

L'impostazione di configurazioni di garbage collection personalizzate con `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions` genera un errore di avvio del driver o dell'executor con Amazon EMR 6.1 a causa di una configurazione di garbage collection in conflitto con Amazon EMR 6.1.0. Per Amazon EMR 6.1.0, la configurazione predefinita per la rimozione di oggetti inutili (garbage collection) viene impostata tramite `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Questa configurazione si applica solo ad Amazon EMR 6.1.0. Le opzioni JVM non correlate alla garbage collection, ad esempio quelle per la configurazione della registrazione (`-verbose:class`) possono comunque essere impostate tramite `extraJavaOptions`. Per ulteriori informazioni, consulta [Proprietà dell'applicazione Spark](#).

Utilizzo di `maximizeResourceAllocation`

Se desideri configurare gli executor per utilizzare la quantità massima delle risorse possibili in ogni nodo di un cluster, imposta `maximizeResourceAllocation` su `true` nella tua classificazione di configurazione `spark.maximizeResourceAllocation` è specifico per Amazon EMR. Quando

abilitati `maximizeResourceAllocation`, EMR calcola le risorse di elaborazione e memoria massime per un executor su un'istanza nel gruppo di istanze principale. In seguito imposta le impostazioni `spark-defaults` corrispondenti in base ai valori massimi calcolati.

Note

Non è consigliabile utilizzare l'opzione `maximizeResourceAllocation` su cluster con altre applicazioni distribuite come HBase. Amazon EMR utilizza configurazioni YARN personalizzate per applicazioni distribuite, che possono entrare in conflitto con `maximizeResourceAllocation` e causare un funzionamento non corretto delle applicazioni Spark.

Di seguito è riportato un esempio di classificazione della configurazione Spark con `maximizeResourceAllocation` impostato su `true`.

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

Impostazioni configurate in `spark-defaults` quando `maximizeResourceAllocation` è abilitata

Impostazione	Descrizione	Valore
<code>spark.default.parallelism</code>	Numero predefinito di partizioni in RDD restituito da trasformazioni come <code>join</code> , <code>reduceByKey</code> e <code>parallelize</code> quando non impostato dall'utente.	2X numero di core CPU disponibili per i container YARN.
<code>spark.driver.memory</code>	Quantità di memoria da utilizzare per il processo di driver, durante il quale	L'impostazione è configurata in base ai tipi di istanza nel cluster. Tuttavia, dato che

Impostazione	Descrizione	Valore
	SparkContext è inizializzato, ad esempio 1g, 2g.	L'applicazione del driver di Spark può essere eseguita sull'istanza primaria o su una delle istanze core (ad esempio, rispettivamente in modalità cluster e client YARN), l'impostazione è definita in base al tipo di istanza più piccolo in questi due gruppi di istanze.
spark.executor.memory	Quantità di memoria da utilizzare per processo esecutore, ad esempio 1g, 2g	L'impostazione è configurata in base ai tipi di istanza principale e di attività nel cluster.
spark.executor.cores	Il numero di core da utilizzare in ogni esecutore.	L'impostazione è configurata in base ai tipi di istanza principale e di attività nel cluster.
spark.executor.instances	Il numero di esecutori.	L'impostazione è configurata in base ai tipi di istanza principale e di attività nel cluster. Da impostare, a meno che <code>spark.dynamicAllocation.enabled</code> sia esplicitamente impostato su <code>true</code> nello stesso tempo.

Configurazione del comportamento di disattivazione di nodi

Quando si utilizza Amazon EMR rilascio 5.9.0 o successivi, Spark su Amazon EMR include un set di caratteristiche che consentono a Spark di gestire normalmente la terminazione dei nodi dovuta a un ridimensionamento manuale o alla richiesta di una policy di dimensionamento automatico.

Amazon EMR implementa un meccanismo di elenco rifiutati in Spark basato sul meccanismo di disattivazione di YARN. Questo meccanismo contribuisce a garantire che nessuna nuova attività sia programmata su un nodo che sarà disattivato, consentendo nel contempo il completamento delle attività in esecuzione. Sono inoltre disponibili caratteristiche che consentono un recupero più rapido dei processi Spark in caso di perdita di blocchi shuffle durante la terminazione di un nodo. Il processo di ricalcolo viene attivato prima e ottimizzato per essere realizzato più rapidamente con un minor numero di tentativi di fase; inoltre, è possibile evitare errori nei processi dovuti a problemi di recupero causati da blocchi shuffle mancanti.

Important

Ad Amazon EMR rilascio 5.11.0 è stata aggiunta l'impostazione `spark.decommissioning.timeout.threshold` per migliorare la resilienza di Spark quando si utilizzano istanze spot. Nei rilasci precedenti, quando un nodo utilizza un'istanza spot e l'istanza viene terminata a causa del prezzo dell'offerta, è possibile che Spark non riesca a gestire normalmente la terminazione. I processi possono non riuscire e i ricalcoli shuffle possono richiedere una notevole quantità di tempo. Per questo motivo, consigliamo di utilizzare il rilascio 5.11.0 o successivi in caso di impiego delle istanze spot.

Impostazioni di disattivazione di nodi Spark

Impostazione	Descrizione	Valore predefinito
<code>spark.blacklist.decommissioning.enabled</code>	Quando il valore è <code>true</code> , Spark elenca come negati i nodi il cui stato è <code>decommissioning</code> in YARN. Spark non pianifica nuove attività su esecutori attivi su quel nodo, ma consente il completamento delle attività già in esecuzione.	<code>true</code>
<code>spark.blacklist.decommissioning.timeout</code>	Il periodo di tempo durante il quale un nodo il cui stato è <code>decommissioning</code> rimane elencato come negato Per	1h

Impostazione	Descrizione	Valore predefinito
	<p>impostazione predefinita, questo valore è impostato su un'ora, che è anche l'impostazione predefinita di <code>yarn.resourcemanager.decommissioning.timeout</code>. Per garantire che un nodo sia elencato come negato per l'intero periodo di disattivazione, impostare un valore uguale o superiore a <code>yarn.resourcemanager.decommissioning.timeout</code>. Alla scadenza del timeout di disattivazione, lo stato del nodo diventa <code>decommissioned</code> e Amazon EMR può terminare l'istanza EC2 del nodo. Se una o più attività sono ancora in esecuzione dopo la scadenza del timeout, sono perse oppure terminate e riprogrammate su esecutori attivi su altri nodi.</p>	

Impostazione	Descrizione	Valore predefinito
<code>spark.decommissioning.timeout.threshold</code>	Disponibile in Amazon EMR rilascio 5.11.0 o successivi. Specificata in secondi. Quando lo stato del nodo diventa "decommissioning" (in disattivazione), se l'host viene disattivato entro un periodo di tempo uguale o inferiore a questo valore, Amazon EMR elenca il nodo come negato e pulisce lo stato dell'host (come specificato da <code>spark.resourceManager.cleanupExpiredHost</code>) senza attendere che lo stato del nodo diventi disattivato. Ciò consente a Spark di gestire meglio le terminazioni delle istanze Spot in quanto la disattivazione di tali istanze avviene entro un timeout di 20 secondi indipendentemente dal valore di <code>yarn.resourcemanager.decommissioning.timeout</code> , che potrebbe non fornire agli altri nodi tempo sufficiente per leggere i file di shuffle.	20s

Impostazione	Descrizione	Valore predefinito
<code>spark.resourceManager.cleanupExpiredHost</code>	Quando impostato su <code>true</code> , Spark annulla la registrazione di tutti i dati memorizzati nella cache e dei blocchi shuffle archiviati in esecutori sui nodi il cui stato è <code>decommissioned</code> . Questa operazione accelera il processo di ripristino.	<code>true</code>
<code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code>	Quando impostata su <code>true</code> , consente di evitare errori nelle fasi Spark e quindi nel processo dovuti a un numero elevato di recuperi non riusciti da nodi disattivati. I recuperi di blocchi shuffle non riusciti da un nodo il cui stato è <code>decommissioned</code> non sono contabilizzati nel numero massimo di errori di recupero consecutivi.	<code>true</code>

Variabile di ambiente ThriftServer Spark

Spark imposta la variabile di ambiente relativa alla porta del server Hive Thrift, `HIVE_SERVER2_THRIFT_PORT`, su 10001.

Modifica delle impostazioni di default di Spark

Puoi modificare le impostazioni di default in `spark-defaults.conf` utilizzando la classificazione di configurazione `spark-defaults` o l'impostazione `maximizeResourceAllocation` nella classificazione di configurazione `spark`.

Le procedure seguenti mostrano come modificare le impostazioni utilizzando la CLI o la console.

Creazione di un cluster con spark.executor.memory impostato su 2g utilizzando la CLI

- Crea un cluster con Spark installato e spark.executor.memory impostato su 2g, utilizzando il comando seguente che fa riferimento a un file, myConfig.json, archiviato in Amazon S3.

```
aws emr create-cluster --release-label emr-5.36.1 --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

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

myConfig.json:

```
[  
  {  
    "Classification": "spark-defaults",  
    "Properties": {  
      "spark.executor.memory": "2G"  
    }  
  }  
]
```

Creazione di un cluster con spark.executor.memory impostato su 2g mediante la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Scegliere Spark.
4. In Edit software settings (Modifica impostazioni software), lasciare l'opzione Enter configuration (Immetti configurazione) selezionata e immettere la seguente configurazione:

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Seleziona le altre opzioni, scegli e in seguito Create cluster (Crea cluster).

Per impostare `maximizeResourceAllocation`

- Crea un cluster con Spark installato e `maximizeResourceAllocation` impostato su `true` (vero) mediante la AWS CLI, facendo riferimento a un file, `myConfig.json`, archiviato in Amazon S3.

```
aws emr create-cluster --release-label emr-5.36.1 --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

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

`myConfig.json`:

```
[  
  {  
    "Classification": "spark",  
    "Properties": {  
      "maximizeResourceAllocation": "true"  
    }  
  }  
]
```

Note

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze

in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Migrazione da Apache Log4j 1.x a Log4j 2.x

Le versioni [Apache Spark](#) 3.2.x e precedenti utilizzano la versione precedente di Apache Log4j 1.x e il file `log4j.properties` per configurare Log4j nei processi Spark. Le versioni 3.3.0 e successive di Apache Spark utilizzano Apache Log4j 2.x e il file `log4j2.properties` per configurare Log4j nei processi Spark.

Se hai configurato Apache Spark Log4j utilizzando una versione Amazon EMR precedente alla 6.8.0, devi rimuovere la classificazione della configurazione `spark-log4j legacy` e migrazione alla classificazione della configurazione `spark-log4j2` e formato delle chiavi prima di poter effettuare l'aggiornamento ad Amazon EMR 6.8.0 o versione successiva. La classificazione `spark-log4j legacy` causa il fallimento della creazione del cluster con un errore `ValidationException` nel rilascio di Amazon EMR versione 6.8.0 e successive. Non ti verrà addebitato alcun costo per un guasto relativo all'incompatibilità di Log4j, ma devi rimuovere la classificazione della configurazione `spark-log4j` passata per continuare.

Per ulteriori informazioni sulla migrazione da Apache Log4j 1.x a Log4j 2.x, consulta [Guida alla migrazione per Apache Log4j](#) e il [Modello Spark Log4j 2](#) su GitHub.

Note

Con Amazon EMR, Apache Spark utilizza un file `log4j2.properties` anziché il file `.xml` descritto nella [Guida alla migrazione per Apache Log4j](#). Inoltre, non è consigliabile utilizzare il metodo `bridge` Log4j 1.x per la conversione in Log4j 2.x.

Ottimizzazione delle prestazioni di Spark

Amazon EMR offre diverse caratteristiche di ottimizzazione delle prestazioni per Spark. In questo argomento viene descritta in dettaglio ogni caratteristica di ottimizzazione.

Per ulteriori informazioni su come impostare la configurazione Spark, consulta [Configurazione di Spark](#).

Esecuzione adattiva delle query

L'esecuzione adattiva delle query è un framework per riottimizzare i piani di query in base alle statistiche di runtime. A partire da Amazon EMR 5.30.0, le seguenti ottimizzazioni di esecuzione adattiva delle query di Apache Spark 3 sono disponibili sul runtime Apache EMR per Spark 2.

- Conversione adattiva di join
- Unione adattiva delle partizioni shuffle

Conversione adattiva di join

La conversione adattiva dei join migliora le prestazioni delle query convertendo le operazioni di sort-merge-join in operazioni di broadcast-hash-join in base alle dimensioni di runtime degli stadi delle query. Le operazioni di broadcast-hash-join tendono a funzionare meglio quando un lato del join è abbastanza piccolo da trasmettere in modo efficiente il suo output su tutti gli executor, evitando così la necessità di mischiare lo scambio e ordinare entrambi i lati del join. La conversione adattiva di join amplia la gamma di casi in cui Spark esegue automaticamente le operazioni broadcast-hash-join.

Questa caratteristica viene attivata per impostazione predefinita. Può essere disabilitata impostando `spark.sql.adaptive.enabled` su `false`, operazione che disabilita anche il framework di esecuzione adattiva delle query. Spark decide di convertire un sort-merge-join in un broadcast-hash-join quando la statistica delle dimensioni del runtime di uno dei lati di join non supera `spark.sql.autoBroadcastJoinThreshold`, che per impostazione predefinita corrisponde a 10.485.760 byte (10 MiB).

Unione adattiva delle partizioni shuffle

L'unione adattiva delle partizioni shuffle migliora le prestazioni delle query unendo piccole partizioni shuffle contigue per evitare il sovraccarico dato da un numero eccessivo di attività minori. Ciò consente di configurare un numero più elevato di partizioni shuffle iniziali che poi vengono ridotte in fase di runtime a una dimensione mirata, migliorando le possibilità di avere partizioni shuffle distribuite in modo più uniforme.

Questa funzionalità è attivata per impostazione predefinita, a meno che `spark.sql.shuffle.partitions` non sia impostato in modo esplicito. Può essere abilitata impostando `spark.sql.adaptive.coalescePartitions.enabled` su `true`. Sia il numero iniziale di partizioni shuffle che la dimensione delle partizioni di destinazione possono essere regolati utilizzando rispettivamente le

proprietà `spark.sql.adaptive.coalescePartitions.minPartitionNum` e `spark.sql.adaptive.advisoryPartitionSizeInBytes`. Per ulteriori dettagli sulle proprietà Spark correlate a questa caratteristica, consulta la tabella riportata di seguito.

Proprietà della partizione di unione adattiva Spark

Proprietà	Valore predefinito	Descrizione
<code>spark.sql.adaptive.coalescePartitions.enabled</code>	true, a meno che <code>spark.sql.shuffle.partition</code> non sia impostato in modo esplicito	Quando true e <code>spark.sql.adaptive.enabled</code> sono impostati su true, Spark unisce le partizioni shuffle contigue in base alla dimensione di destinazione (specificata da <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>), in modo da evitare l'accumulo di attività minori.
<code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>	64 MB	La dimensione degli advisory in byte della partizione shuffle durante l'unione. Questa configurazione ha effetto solo quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions.enabled</code> sono entrambi true.
<code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code>	25	Il numero minimo di partizioni shuffle dopo l'unione. Questa configurazione ha effetto solo quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions</code>

Proprietà	Valore predefinito	Descrizione
		<code>s.enabled</code> sono entrambi <code>true</code> .
<code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code>	1000	Il numero iniziale di partizioni shuffle prima dell'unione. Questa configurazione ha effetto solo quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions.enabled</code> sono entrambi <code>true</code> .

Eliminazione delle partizioni dinamiche

L'eliminazione delle partizioni dinamiche migliora le prestazioni del processo selezionando accuratamente partizioni specifiche all'interno di una tabella che devono essere lette ed elaborate per una query specifica. La riduzione della quantità di dati letti ed elaborati determina un risparmio di tempo significativo nell'esecuzione del processo. Questa caratteristica viene attivata per impostazione predefinita con Amazon EMR 5.26.0. Con Amazon EMR 5.24.0 e 5.25.0 puoi abilitare questa caratteristica impostando le proprietà Spark `spark.sql.dynamicPartitionPruning.enabled` da Spark o durante la creazione di cluster.

Proprietà della partizione di eliminazione dinamica di partizioni Spark

Proprietà	Valore predefinito	Descrizione
<code>spark.sql.dynamicPartitionPruning.enabled</code>	<code>true</code>	Quando è <code>true</code> , abilitare l'eliminazione dinamica delle partizioni.
<code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code>	<code>true</code>	Quando è <code>true</code> , Spark esegue un controllo difensivo prima dell'esecuzione delle query per garantire che il riutilizzo degli scambi

Proprietà	Valore predefinito	Descrizione
		<p>broadcast nei filtri di eliminazione dinamica non venga interrotto da regole di preparazione successive, ad esempio regole colonnari definite dall'utente. Quando il riutilizzo è interrotto e questa configurazione è <code>true</code>, Spark rimuove i filtri di eliminazione dinamici interessati per evitare problemi di prestazioni e correttezza. I problemi di correttezza possono sorgere quando lo scambio broadcast del filtro di eliminazione dinamica produce risultati diversi e incoerenti dallo scambio broadcast dell'operazione di join corrispondente. L'impostazione di questa configurazione su <code>false</code> deve essere eseguita con cautela, in quanto consente di aggirare gli scenari, ad esempio quando il riutilizzo viene interrotto da regole colonnari definite dall'utente. Quando Adaptive Query Execution (Esecuzione adattiva delle query) è abilitata, il riutilizzo broadcast viene sempre applicato.</p>

Questa ottimizzazione migliora le funzionalità esistenti di Spark 2.4.2, che supportano solo la distribuzione dei predicati che possono essere risolti in fase di pianificazione.

Di seguito sono elencati alcuni esempi di distribuzione di predicati statici in Spark 2.4.2.

```
partition_col = 5

partition_col IN (1,3,5)

partition_col between 1 and 3

partition_col = 1 + 3
```

L'eliminazione delle partizioni dinamiche consente al motore Spark di dedurre dinamicamente in fase di runtime quali partizioni devono essere lette e quali possono essere eliminate. Ad esempio, la seguente query comporta due tabelle: la tabella `store_sales` contenente tutte le vendite totali di tutti gli store, partizionata in base alla regione, e la tabella `store_regions` contenente una mappatura di regioni per ogni paese. Le tabelle contengono i dati relativi agli store distribuiti in tutto il mondo, ma vengono eseguite query solo per il Nord America.

```
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'
```

Senza l'eliminazione delle partizioni dinamiche, questa query leggerà tutte le regioni prima di filtrare il sottoinsieme di regioni che corrispondono ai risultati della subquery. Con l'eliminazione delle partizioni dinamiche, questa query legge ed elabora solo le partizioni per le regioni restituite nella subquery. Questo consente di risparmiare tempo e risorse leggendo meno dati dallo storage ed elaborando meno record.

Appiattimento delle subquery scalari

Questa ottimizzazione migliora le prestazioni di query che dispongono di subquery scalari sulla stessa tabella. Questa caratteristica viene attivata per impostazione predefinita con Amazon EMR 5.26.0. Con Amazon EMR 5.24.0 e 5.25.0, puoi abilitarla impostando la proprietà Spark `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` da Spark o durante la creazione di cluster. Quando questa proprietà è impostata su `true`, l'ottimizzatore di query appiattisce le subquery scalari accumulate che utilizzano la stessa relazione se possibile. Le subquery scalari vengono appiattite distribuendo gli eventuali predicati presenti nella subquery nelle

funzioni di aggregazione e quindi eseguendo una aggregazione, con tutte le funzioni di aggregazione, per relazione.

Di seguito è riportato un esempio di query che sfrutta questa ottimizzazione.

```
select (select avg(age) from students          /* Subquery 1 */
       where age between 5 and 10) as group1,
       (select avg(age) from students          /* Subquery 2 */
       where age between 10 and 15) as group2,
       (select avg(age) from students          /* Subquery 3 */
       where age between 15 and 20) as group3
```

L'ottimizzazione riscrive la query precedente come:

```
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);
```

Nota che la query riscritta legge la tabella studenti una sola volta e i predicati delle tre subquery vengono distribuiti nella funzione avg.

DISTINCT prima di INTERSECT

Questa ottimizzazione ottimizza i join durante l'utilizzo di INTERSECT. Questa caratteristica viene attivata per impostazione predefinita con Amazon EMR 5.26.0. Con Amazon EMR 5.24.0 e 5.25.0, puoi abilitarla impostando la proprietà Spark `spark.sql.optimizer.distinctBeforeIntersect.enabled` da Spark o durante la creazione di cluster. Le query che utilizzano INTERSECT vengono automaticamente convertite per utilizzare un join left-semi. Quando questa proprietà è impostata su VERO, l'ottimizzatore di query distribuisce l'operatore DISTINCT nei figli di INTERSECT se rileva che l'operatore DISTINCT può rendere il join left-semi un BroadcastHashJoin anziché un SortMergeJoin.

Di seguito è riportato un esempio di query che sfrutta questa ottimizzazione.

```
(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)
```

Senza l'abilitazione di questa proprietà `spark.sql.optimizer.distinctBeforeIntersect.enabled`, la query sarà riscritta come segue.

```
select distinct brand from
  (select item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

Quando abiliti questa proprietà `spark.sql.optimizer.distinctBeforeIntersect.enabled`, la query sarà riscritta come segue.

```
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
```

Join del filtro Bloom

Questa ottimizzazione può migliorare le prestazioni di alcuni join mediante il prefiltraggio di un lato di un join utilizzando un filtro [Bloom](#) generato dai valori dell'altro lato del join. Questa caratteristica viene attivata per impostazione predefinita con Amazon EMR 5.26.0. Con Amazon EMR 5.25.0 puoi abilitare questa caratteristica impostando la proprietà `spark.sql.bloomFilterJoin.enabled` di Spark su `true` da Spark o durante la creazione di cluster.

Di seguito è riportato un esempio di query che può trarre vantaggio da un filtro Bloom.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

Quando questa caratteristica è abilitata, il filtro Bloom viene creato da tutti gli ID degli elementi la cui categoria è presente nel set delle categorie su cui viene eseguita la query. Durante la scansione della tabella SALES, il filtro Bloom viene utilizzato per determinare quali sono le vendite relative a elementi che non sono presenti nel set definito dal filtro Bloom. Pertanto, tali vendite identificate possono essere escluse dal filtro il prima possibile.

Riordinamento ottimizzato dei join

Questa ottimizzazione può migliorare le prestazioni delle query riordinando i join che includono tabelle con filtri. Questa caratteristica viene attivata per impostazione predefinita con Amazon EMR 5.26.0. Con Amazon EMR 5.25.0, è possibile abilitare questa caratteristica impostando il parametro di configurazione Spark `spark.sql.optimizer.sizeBasedJoinReorder.enabled` su "true" (vero). Il comportamento predefinito in Spark è di eseguire il join delle tabelle da sinistra a destra, come elencato nella query. Questa strategia potrebbe non dare l'opportunità di eseguire prima i join più piccoli con i filtri, in modo da trarre vantaggio per join più onerosi in un momento successivo.

La seguente query di esempio segnala tutti gli elementi restituiti da tutti gli store in un paese. Senza il riordinamento ottimizzato dei join, Spark esegue prima il join delle due tabelle di grandi dimensioni, `store_sales` e `store_returns`, quindi ne esegue il join con `store` e infine con `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'
```

Con il riordinamento ottimizzato dei join, Spark esegue il join di `store_sales` con `store` per prima cosa poiché `store` ha un filtro ed è inferiore a `store_returns` e a `broadcastable`. Quindi Spark esegue il join con `store_returns` e infine con `item`. Se `item` avesse avuto un filtro e fosse stato trasmissibile, sarebbe stato qualificato per il riordinamento, con il risultato di un join di `store_sales` con `store`, quindi con `item` e infine con `store_returns`.

Caching dei frammenti dei risultati Spark

Amazon EMR 6.6.0 e versioni successive includono la funzione opzionale Caching dei frammenti dei risultati Spark che memorizza automaticamente nella cache i frammenti dei risultati. Questi frammenti di risultati sono parti dei risultati di sottoalberi di query archiviati in un bucket Amazon S3 di tua scelta. I frammenti dei risultati della query memorizzati vengono riutilizzati nelle successive esecuzioni di query, rilasciando query più veloci.

Il Caching dei frammenti dei risultati funziona analizzando le query Spark SQL e memorizzando nella cache i frammenti di risultati idonei nella posizione S3 specificata. Nelle successive esecuzioni di query, i frammenti dei risultati della query utilizzabili vengono rilevati automaticamente e recuperati da S3. Il Caching dei frammenti di risultato è diverso dal Caching dei set di risultati, in cui le query successive devono corrispondere esattamente alla query originale per restituire risultati dalla cache. Se utilizzato per query che si rivolgono ripetutamente a un sottoinsieme statico di dati, la memorizzazione nella cache dei frammenti di risultato accelera notevolmente le prestazioni.

Si consideri la seguente query, che conta gli ordini fino all'anno 2022:

```
select
  l_returnflag,
  l_linestatus,
  count(*) as count_order
from
  lineitem
where
  l_shipdate <= current_date
  and year(l_shipdate) == '2022'
group by
  l_returnflag,
  l_linestatus
```

Con il passare del tempo, questa query deve essere eseguita ogni giorno per segnalare le vendite totali per l'anno. Senza Caching dei frammenti dei risultati, i risultati per tutti i giorni dell'anno dovranno essere ricalcolati ogni giorno. La query diventerà più lenta nel tempo e sarà più lenta alla fine dell'anno, quando tutti i 365 giorni di risultati dovranno essere ricalcolati.

Quando si attiva Caching dei frammenti dei risultati, si utilizzano i risultati di tutti i giorni precedenti dell'anno dalla cache. Ogni giorno, la funzione deve ricalcolare solo un giorno di risultati. Dopo che la funzione ha calcolato il frammento del risultato, questa memorizza nella cache il frammento. Di conseguenza, i tempi di query abilitati per la cache sono rapidi e rimangono costanti per ogni query successiva.

Abilitazione del Caching dei frammenti dei risultati Spark

Per abilitare la Caching dei frammenti dei risultati di Spark, segui la procedura seguente:

1. Crea un bucket di cache in Amazon S3 e autorizza l'accesso in lettura/scrittura per EMRFS. Per ulteriori informazioni, consulta [Autorizzazione di accesso ai dati di EMRFS in Amazon S3](#).

2. Imposta la configurazione di EMR Spark per abilitare la funzione.

```
spark.subResultCache.enabled = true
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Abilita la gestione del ciclo di vita S3 per il bucket per pulire automaticamente i file della cache.
4. Facoltativamente, configura le proprietà `reductionRationThreshold` e `maxBufferSize` per ottimizzare ulteriormente la funzione.

```
spark.sql.subResultCache.reductionRatioThreshold
spark.sql.subResultCache.maxBufferSize
```

Considerazioni su quando utilizzare il Caching dei frammenti

Il risparmio sui costi quando si utilizzano risultati già memorizzati nella cache in Amazon S3 anziché ricalcolarli aumenta con il numero di volte in cui è possibile utilizzare gli stessi risultati memorizzati nella cache. Le query con scansioni di tabelle di grandi dimensioni seguite da filtri o aggregazioni hash che riducono la dimensione del risultato di un fattore di almeno 8 (vale a dire, un rapporto di almeno 8:1 in dimensione di `size:results`) trarranno i maggiori vantaggi da questa funzione. Maggiore è il rapporto di riduzione tra input e risultati, maggiore è il rapporto costi-benefici. Ne trarranno vantaggio anche le query con rapporti di riduzione inferiori, ma che contengono passaggi di calcolo costosi tra la scansione della tabella e il filtro o le aggregazioni, purché il costo per produrre i risultati sia maggiore del costo per recuperarli da Amazon S3. Per impostazione predefinita, Caching dei frammenti dei risultati ha effetto solo quando rileva che un rapporto di riduzione sarà almeno di 8:1.

Quando le tue query riutilizzano ripetutamente i risultati memorizzati nella cache, i vantaggi di questa funzione sono maggiori. Le query a finestra continua e incrementali sono buoni esempi. Ad esempio, una query a finestra continua di 30 giorni che è già stata eseguita per 29 giorni, avrebbe solo bisogno di estrarre 1/30 dei dati di destinazione dalla sua origine di input originale e utilizzerebbe frammenti di risultati memorizzati nella cache per i 29 giorni precedenti. Una query a finestra incrementale ne trarrebbe un vantaggio ancora maggiore, poiché l'inizio della finestra rimane fisso: ad ogni invocazione della query, una percentuale minore dell'elaborazione richiederà la lettura dall'origine di input.

Di seguito sono riportate ulteriori considerazioni quando si utilizza il Caching dei frammenti dei risultati:

- Le query che non hanno come target gli stessi dati con gli stessi frammenti di query avranno una percentuale di riscontri nella cache bassa, quindi non trarranno vantaggio da questa funzione.
- Le query con rapporti di riduzione bassi che non contengono passaggi di calcolo costosi porteranno a risultati memorizzati nella cache che sono all'incirca tanto costosi da leggere quanto lo erano per l'elaborazione iniziale.
- La prima query mostrerà sempre una regressione minore a causa del costo di scrittura nella cache.
- La funzione di memorizzazione nella cache di frammenti risultanti funziona esclusivamente con i file Parquet. Altri formati di file non sono supportati.
- I buffer della funzione di memorizzazione nella cache di frammenti risultanti tenteranno di memorizzare nella cache solo scansioni con dimensioni di suddivisione dei file pari o superiori a 128 MB. Con la configurazione Spark predefinita, il Caching dei frammenti dei risultati sarà disabilitato se la dimensione di scansione (dimensione totale di tutti i file sottoposti a scansione) divisa per il numero di core esecutori è inferiore a 128 MB. Quando viene impostata una delle configurazioni Spark elencate di seguito, la dimensione della divisione del file sarà:

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- `spark.sql.leafNodeDefaultParallelism` (il valore predefinito è `spark.default.parallelism`)
- `Spark.sql.files.minPartitionNum` (il valore predefinito è `spark.sql.leafNodeDefaultParallelism`)
- `spark.sql.files.openCostInBytes`
- `spark.sql.files.maxPartitionBytes`
- La funzione di memorizzazione nella cache di frammenti risultanti memorizza nella cache la granularità della partizione RDD. Il rapporto di riduzione precedentemente descritto che per impostazione predefinita è 8:1 viene valutato per partizione RDD. I carichi di lavoro con rapporti di riduzione per RDD maggiori o inferiori a 8:1 possono riscontrare vantaggi in termini di prestazioni inferiori rispetto ai carichi di lavoro con rapporti di riduzione per RDD costantemente inferiori a 8:1.
- La funzione di memorizzazione nella cache di frammenti risultanti utilizza un buffer di scrittura da 16 MB per impostazione predefinita per ogni partizione RDD che viene memorizzata nella cache. Se più di 16 MB verranno memorizzati nella cache per partizione RDD, il costo per determinare l'impossibilità di una scrittura può comportare una regressione delle prestazioni.
- Mentre, per impostazione predefinita, il Caching dei frammenti dei risultati non tenterà di memorizzare nella cache i risultati della partizione RDD con un rapporto di riduzione inferiore a 8:1 e limiterà il suo buffer di scrittura a 16 MB; entrambi questi valori sono regolabili attraverso le seguenti configurazioni:

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Più cluster che utilizzano lo stesso rilascio di EMR possono condividere la stessa posizione della cache. Per garantire la correttezza dei risultati, la memorizzazione nella cache di frammenti risultanti non utilizzerà i risultati della cache scritti da rilasci differenti di Amazon EMR.
- Il Caching dei frammenti dei risultati verrà disabilitato automaticamente per i casi d'uso di Spark Streaming o quando viene utilizzato RecordServer, Apache Ranger o AWS Lake Formation.
- Le operazioni di lettura/scrittura nella cache dei frammenti del risultato utilizzano bucket EMRFS e Amazon S3. La crittografia CSE/ SSE S3/ SSE KMS è supportata.

Utilizzare l'Nvidia Spark-RAPIDS Accelerator per Spark

Con Amazon EMR rilascio 6.2.0 e successivi, puoi utilizzare il plug-in dell'acceleratore [RAPIDS](#) di Nvidia per Apache Spark per accelerare Spark utilizzando tipi di istanza Graphics Processing Unit (GPU) di EC2. L'Accelerator Rapids esegue l'accelerazione via GPU delle pipeline Data Science di Apache Spark 3.0 senza applicare modifiche al codice e velocizza l'elaborazione dei dati e l'addestramento dei modelli, riducendo notevolmente i costi dell'infrastruttura.

Le sezioni seguenti guidano l'utente nella configurazione del cluster EMR per l'utilizzo del plug-in SPARK-Rapids per Spark.

Scelta dei tipi di istanza

Per utilizzare il plugin Nvidia Spark-RAPIDS per Spark, i gruppi di istanze principali e attività devono utilizzare i tipi di istanza GPU EC2 che soddisfano i [Requisiti hardware](#) di Spark-RAPIDS. Per visualizzare un elenco completo dei tipi di istanza GPU supportati da EMR, consulta i [Tipi di istanze supportati](#) nella Guida alla gestione di Amazon EMR. Il tipo di istanza per il gruppo di istanze primarie può essere GPU o non GPU, ma i tipi di istanza ARM non sono supportati.

Impostazione delle configurazioni delle applicazioni per il cluster

1. Abilitare Amazon EMR per installare i plug-in nel nuovo cluster

Per installare i plug-in, è necessaria la seguente configurazione durante la creazione del cluster:

```
{
  "Classification": "spark",
```

```

"Properties":{
  "enableSparkRapids":"true"
}
}

```

2. Configurazione di YARN per utilizzare la GPU

Per ulteriori informazioni sull'utilizzo della GPU su YARN, consulta [Using GPU on YARN \(Utilizzo della GPU su YARN\)](#) nella documentazione di Apache Hadoop. Segue una configurazione di esempio:

```

{
  "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. Configurazione di Spark per utilizzare RAPIDS

Seguono le configurazioni necessarie per consentire a Spark di utilizzare il plug-in RAPIDS:

```

{
  "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 library](#) nella documentazione di XGBoost è disponibile anche quando il plug-in Spark RAPIDS è abilitato sul cluster. Puoi utilizzare la seguente configurazione per integrare XGBoost con il tuo processo Spark:

```

{
  "Classification":"spark-defaults",
  "Properties":{
    "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
  }
}

```

Per ulteriori configurazioni Spark che è possibile utilizzare per regolare un cluster EMR accelerato da GPU, fai riferimento alla [Rapids Accelerator for Apache Spark tuning guide \(Guida alla regolazione dell'Accelerator Rapids per Apache Spark\)](#) nella documentazione di Nvidia.github.io.

4. Configurazione del pianificatore di capacità YARN

`DominantResourceCalculator` deve essere configurato per abilitare la pianificazione e l'isolamento della GPU. Per ulteriori informazioni, consulta [Using GPU on YARN \(Utilizzo della GPU su YARN\)](#) nella documentazione di Apache Hadoop.

```
{
  "Classification": "capacity-scheduler",
  "Properties": {
    "yarn.scheduler.capacity.resource-
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
  }
}
```

5. Creazione di un file JSON per includere tutte le configurazioni

È possibile creare un file JSON che contiene la configurazione per l'utilizzo del plug-in RAPIDS per il cluster Spark. È possibile fornire il file in un secondo momento all'avvio del cluster.

Il file può essere memorizzato localmente o su S3. Per ulteriori informazioni su come fornire le configurazioni delle applicazioni per i cluster, consulta [Configurazione delle applicazioni](#).

Di seguito è riportato un esempio di file denominato `my-configurations.json`. Puoi utilizzarlo come modello per iniziare a creare le tue configurazioni.

```
[
  {
    "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"
}
}
]
```

Aggiunta di un'operazione di bootstrap per il cluster

Per utilizzare YARN sulla GPU, è necessario aprire le autorizzazioni cgroups a YARN sul cluster, operazione che può essere eseguita utilizzando uno script di operazione di bootstrap EMR.

Per ulteriori informazioni su come fornire script dell'operazione di bootstrap durante la creazione del cluster, consulta [Nozioni di base sulle operazioni di bootstrap](#) nella Guida alla gestione di Amazon EMR.

Segue uno script di esempio denominato `my-bootstrap-action.sh`:

```
#!/bin/bash

set -ex

sudo chmod a+rxw -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rxw -R /sys/fs/cgroup/devices
```

Avvio del cluster

L'ultimo passaggio consiste nell'avviare il cluster con le configurazioni del cluster menzionate sopra. Segue un esempio del comando per avviare un cluster tramite la CLI EMR:

```
aws emr create-cluster \
--release-label emr-6.2.0 \
--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
```

Accesso alla shell Spark

La shell Spark è basata sul modello Scala REPL (Read-Eval-Print-Loop) e consente di creare programmi Spark in modo interattivo e di inviare lavoro al framework. Puoi accedere alla shell (interprete di comandi) Spark tramite connessione al nodo primario con SSH e richiamando `spark-shell`. Per informazioni su come connettersi al nodo primario, consulta la sezione [Connect to the primary node using SSH](#) (Connessione al nodo primario tramite SSH) nella Guida per la gestione di Amazon EMR. Gli esempi seguenti utilizzano i log di accesso di Apache HTTP Server archiviati in Amazon S3.

Note

Il bucket utilizzato in questi esempi è disponibile per i client che possono accedere alla regione Stati Uniti orientali (Virginia settentrionale).

Per impostazione predefinita, la shell Spark crea un oggetto [SparkContext](#) denominato `sc`. È possibile usare questo contesto se è obbligatorio all'interno della REPL. `sqlContext` è disponibile anche nella shell ed è un [HiveContext](#).

Example Utilizzare la shell Spark per conteggiare le occorrenze di una stringa in un file archiviato in Amazon S3

In questo esempio viene utilizzato `sc` per leggere un `textFile` 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 crea il `textFile` e la [struttura di dati](#) associata. Successivamente, l'esempio conteggia il numero di righe nel file di log con la stringa `"cartoonnetwork.com"`:

```
scala> val linesWithCartoonNetwork = textFile.filter(line =>
  line.contains("cartoonnetwork.com")).count()
```

```
linesWithCartoonNetwork: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at
  filter at <console>:23
<snip>
<Spark program runs>
scala> linesWithCartoonNetwork
res2: Long = 9
```

Example Utilizzare la shell Spark basata su Python per conteggiare le occorrenze di una stringa in un file archiviato in Amazon S3

Spark include anche una shell basata su Python, `pyspark`, che puoi utilizzare per realizzare prototipi di programmi Spark scritti in Python. Come con `spark-shell`, invoca `pyspark` sul nodo primario. Ha inoltre lo stesso oggetto [SparkContext](#).

```
>>> 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 crea il `textFile` e la [struttura di dati](#) associata. Successivamente, l'esempio conteggia il numero di righe nel file di log con la stringa `"cartoonnetwork.com"`.

```
>>> linesWithCartoonNetwork = textfile.filter(lambda line: "cartoonnetwork.com" in
  line).count()
15/06/04 17:12:22 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library from the
  embedded binaries
15/06/04 17:12:22 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo
  library [hadoop-lzo rev EXAMPLE]
15/06/04 17:12:23 INFO fs.EmrFileSystem: Consistency disabled, using
  com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation
<snip>
<Spark program continues>
>>> linesWithCartoonNetwork
9
```

Utilizzare Amazon SageMaker Spark per il machine learning

Quando utilizzi Amazon EMR rilascio 5.11.0 e successivi, il componente `aws-sagemaker-spark-sdk` è installato con Spark. Questo componente installa Amazon SageMaker Spark e le dipendenze associate per l'integrazione di Spark con [Amazon SageMaker](#). Puoi utilizzare Amazon SageMaker

Spark per configurare le pipeline di machine learning (ML) di Spark utilizzando le fasi di Amazon SageMaker. Per ulteriori informazioni, consulta [File Readme di Amazon SageMaker Spark](#) su GitHub e [Utilizzo di Apache Spark con Amazon SageMaker](#) nella Guida per gli sviluppatori di Amazon SageMaker.

Scrittura di un'applicazione Spark

Le applicazioni [Spark](#) possono essere scritte in Scala, Java o Python. Vari esempi di applicazioni Spark sono disponibili nell'argomento [Spark examples \(Esempi Spark\)](#) nella documentazione di Apache Spark. L'esempio di stima del numero Pi è illustrato di seguito nelle tre applicazioni supportate a livello nativo. Puoi anche visualizzare esempi completi in `$SPARK_HOME/examples` e su [GitHub](#). Per ulteriori informazioni su come creare file JAR per Spark, consulta l'argomento [Quick Start \(Avvio rapido\)](#) nella documentazione di Apache Spark.

Scala

Per evitare problemi di compatibilità con Scala, ti consigliamo di utilizzare le dipendenze Spark per la versione corretta di Scala quando compili un'applicazione Spark per un cluster Amazon EMR. La versione Scala che dovresti utilizzare dipende dalla versione di Spark installata nel tuo cluster. Ad esempio, Amazon EMR rilascio 5.30.1 impiega Spark 2.4.5, che è costruito con Scala 2.11. Se il cluster utilizza Amazon EMR rilascio 5.30.1, utilizza le dipendenze Spark per Scala 2.11. Per ulteriori informazioni sulle versioni di Scala utilizzate da Spark, consulta la [Documentazione di Apache Spark](#).

```
package org.apache.spark.examples
import scala.math.random
import org.apache.spark._

/** Computes an approximation to pi */
object SparkPi {
  def main(args: Array[String]) {
    val conf = new SparkConf().setAppName("Spark Pi")
    val spark = new SparkContext(conf)
    val slices = if (args.length > 0) args(0).toInt else 2
    val n = math.min(100000L * slices, Int.MaxValue).toInt // avoid overflow
    val count = spark.parallelize(1 until n, slices).map { i =>
      val x = random * 2 - 1
      val y = random * 2 - 1
      if (x*x + y*y < 1) 1 else 0
    }.reduce(_ + _)
    println("Pi is roughly " + 4.0 * count / n)
  }
}
```

```
    spark.stop()
  }
}
```

Java

```
package org.apache.spark.examples;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;
import org.apache.spark.api.java.function.Function2;

import java.util.ArrayList;
import java.util.List;

/**
 * Computes an approximation to pi
 * Usage: JavaSparkPi [slices]
 */
public final class JavaSparkPi {

    public static void main(String[] args) throws Exception {
        SparkConf sparkConf = new SparkConf().setAppName("JavaSparkPi");
        JavaSparkContext jsc = new JavaSparkContext(sparkConf);

        int slices = (args.length == 1) ? Integer.parseInt(args[0]) : 2;
        int n = 100000 * slices;
        List<Integer> l = new ArrayList<Integer>(n);
        for (int i = 0; i < n; i++) {
            l.add(i);
        }

        JavaRDD<Integer> dataSet = jsc.parallelize(l, slices);

        int count = dataSet.map(new Function<Integer, Integer>() {
            @Override
            public Integer call(Integer integer) {
                double x = Math.random() * 2 - 1;
                double y = Math.random() * 2 - 1;
                return (x * x + y * y < 1) ? 1 : 0;
            }
        });
    }
}
```

```
}).reduce(new Function2<Integer, Integer, Integer>() {
    @Override
    public Integer call(Integer integer, Integer integer2) {
        return integer + integer2;
    }
});

System.out.println("Pi is roughly " + 4.0 * count / n);

jsc.stop();
}
}
```

Python

```
import argparse
import logging
from operator import add
from random import random

from pyspark.sql import SparkSession

logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format="%(levelname)s: %(message)s")

def calculate_pi(partitions, output_uri):
    """
    Calculates pi by testing a large number of random numbers against a unit circle
    inscribed inside a square. The trials are partitioned so they can be run in
    parallel on cluster instances.

    :param partitions: The number of partitions to use for the calculation.
    :param output_uri: The URI where the output is written, typically an Amazon S3
        bucket, such as 's3://example-bucket/pi-calc'.
    """

    def calculate_hit(_):
        x = random() * 2 - 1
        y = random() * 2 - 1
        return 1 if x**2 + y**2 < 1 else 0
```

```
tries = 100000 * partitions
logger.info(
    "Calculating pi with a total of %s tries in %s partitions.", tries, partitions
)
with SparkSession.builder.appName("My PyPi").getOrCreate() as spark:
    hits = (
        spark.sparkContext.parallelize(range(tries), partitions)
        .map(calculate_hit)
        .reduce(add)
    )
    pi = 4.0 * hits / tries
    logger.info("%s tries and %s hits gives pi estimate of %s.", tries, hits, pi)
    if output_uri is not None:
        df = spark.createDataFrame([(tries, hits, pi)], ["tries", "hits", "pi"])
        df.write.mode("overwrite").json(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "--partitions",
        default=2,
        type=int,
        help="The number of parallel partitions to use when calculating pi.",
    )
    parser.add_argument(
        "--output_uri", help="The URI where output is saved, typically an S3 bucket."
    )
    args = parser.parse_args()

    calculate_pi(args.partitions, args.output_uri)
```

Migliorare le prestazioni di Spark con Amazon S3

Amazon EMR offre caratteristiche che aiutano a ottimizzare le prestazioni quando si usa Spark per eseguire query e leggere e scrivere i dati salvati in Amazon S3.

[S3 Select](#) può migliorare le prestazioni delle query per i file CSV e JSON in alcune applicazioni mediante il pushdown dell'elaborazione ad Amazon S3.

Il committer ottimizzato S3 EMRFS è un'alternativa alla classe [OutputCommitter](#), che utilizza la caratteristica di caricamento in più parti EMRFS per migliorare le prestazioni quando si scrivono i file Parquet in Amazon S3 utilizzando Spark SQL, DataFrames e set di dati.

Argomenti

- [Usare S3 Select con Spark per migliorare le prestazioni delle query](#)
- [Utilizzare il committer ottimizzato S3 EMRFS](#)
- [Utilizzo del protocollo di commit ottimizzato per S3 EMRFS](#)
- [Effettuare nuovi tentativi per le richieste Amazon S3 con EMRFS](#)

Usare S3 Select con Spark per migliorare le prestazioni delle query

Con Amazon EMR rilascio 5.17.0 e successivi, puoi utilizzare [S3 Select](#) con Spark su Amazon EMR. S3 Select consente alle applicazioni il recupero di un solo sottoinsieme di dati da un oggetto. Per Amazon EMR, l'attività di calcolo di filtro di set di dati di grandi dimensioni per l'elaborazione viene "trasferita" dal cluster ad Amazon S3; in tal modo è possibile migliorare le prestazioni in alcune applicazioni e ridurre la quantità di dati trasferiti tra Amazon EMR e Amazon S3.

S3 Select è supportata con file CSV e JSON che utilizzano i valori `s3selectCSV` e `s3selectJSON` per specificare il formato dei dati. Per maggiori informazioni ed esempi, consulta [Specificare S3 Select nel codice](#).

S3 Select è adatto alla mia applicazione?

È consigliabile eseguire analisi comparative delle applicazioni con e senza S3 Select per vedere se l'utilizzo di tale caratteristica può risultare idoneo per l'applicazione.

Per determinare se l'applicazione è idonea per l'uso di S3 Select, utilizza le seguenti linee guida:

- La query filtra più di metà del set di dati originale.
- La connessione di rete tra Amazon S3 e il cluster Amazon EMR ha una buona velocità di trasferimento e larghezza di banda disponibile. Amazon S3 non comprime le risposte HTTP, pertanto è probabile che le dimensioni della risposta aumentino per file di input compressi.

Considerazioni e limitazioni

- La crittografia lato server con chiavi di crittografia fornite dal cliente (SSE-C) e la crittografia lato client di Amazon S3 non sono supportate.
- La proprietà `AllowQuotedRecordDelimiters` non è supportata. Se questa proprietà è specificata, la query ha esito negativo.
- Sono supportati solo i file CSV e JSON in formato UTF-8. I file CSV a più righe non sono supportati.
- Sono supportati solo i file non compressi o gzip.
- Le opzioni Spark CSV e JSON quali `nanValue`, `positiveInf`, `negativeInf` e le opzioni relative a record danneggiati (ad esempio, le modalità `failfast` e `dropmalformed`) non sono supportate.
- L'utilizzo di virgole (,) all'interno di decimali non è supportato. Ad esempio, `10,000` non è supportato, mentre è supportato `10000`.
- I caratteri di commento nell'ultima riga non sono supportati.
- Le righe vuote alla fine di un file non vengono elaborate.
- I filtri seguenti non sono trasferiti ad Amazon S3:
 - Funzioni di aggregazione quali `COUNT()` e `SUM()`.
 - Filtri che eseguono il `CAST()` di un attributo. Ad esempio, `CAST(stringColumn as INT) = 1`.
 - Filtri con un attributo che è un oggetto o è complesso. Ad esempio, `intArray[1] = 1`, `objectColumn.objectNumber = 1`.
 - I filtri per i quali il valore non è un valore letterale. Ad esempio, `intColumn1 = intColumn2`
 - Sono supportati solo [tipi di dati supportati da S3 Select](#) con le limitazioni documentate.

Specificare S3 Select nel codice

I seguenti esempi illustrano come specificare S3 Select per CSV utilizzando Scala, SQL, R e PySpark. È possibile utilizzare S3 Select per JSON nello stesso modo. Per un elenco delle opzioni, dei relativi valori predefiniti e le limitazioni, consulta [Opzioni](#).

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")
```

Opzioni

Le seguenti opzioni sono disponibili quando si utilizzano `s3selectCSV` e `s3selectJSON`. Se non specificate, verranno utilizzati i valori predefiniti.

Opzioni con S3selectCSV

Opzione	Default (Predefinito)	Utilizzo
<code>compression</code>	"none"	Indica se viene utilizzata la compressione. "gzip" è l'unica impostazione supportata oltre a "none".

Opzione	Default (Predefinito)	Utilizzo
<code>delimiter</code>	<code>","</code>	Specifica il delimitatore del campo.
<code>quote</code>	<code>'\"'</code>	Specifica il carattere virgolettato. La specifica di una stringa vuota non è supportata e genera un errore di XML danneggiato.
<code>escape</code>	<code>'\\'</code>	Specifica il carattere di escape.
<code>header</code>	<code>"false"</code>	<code>"false"</code> specifica che non è presente alcuna intestazione. <code>"true"</code> specifica che nella prima riga è presente un'intestazione. Sono supportate solo le intestazioni nella prima riga e non sono supportate righe vuote prima di un'intestazione.
<code>comment</code>	<code>"#"</code>	Specifica il carattere per commenti. Non è possibile disabilitare l'indicatore dei commenti. In altre parole, il valore <code>\u0000</code> non è supportato.
<code>nullValue</code>	<code>""</code>	

Opzioni con S3selectJSON

Opzione	Default (Predefinito)	Utilizzo
<code>compression</code>	<code>"none"</code>	Indica se viene utilizzata la compressione. <code>"gzip"</code> è l'unica impostazione supportata oltre a <code>"none"</code> .
<code>multiline</code>	<code>"false"</code>	<code>"false"</code> specifica che JSON è nel formato LINES S3 Select, che significa che ogni riga dei dati di input contiene un singolo oggetto JSON. <code>"true"</code> specifica che JSON è nel formato DOCUMENT S3 Select, che significa che un oggetto JSON può estendersi su più righe dei dati di input.

Utilizzare il committer ottimizzato S3 EMRFS

Il committer ottimizzato S3 EMRFS è un'alternativa all'implementazione di [OutputCommitter](#) che è ottimizzata per scrivere i file in Amazon S3 quando si usa EMRFS. Il committer ottimizzato per S3 EMRFS migliora le prestazioni delle applicazioni evitando l'esecuzione di operazioni di elenchi e di ridenominazione in Amazon S3 durante le fasi di commit di processi e attività. Il committer è disponibile in Amazon EMR rilascio 5.19.0 e successivi ed è abilitato per impostazione predefinita in Amazon EMR rilascio 5.20.0 e successivi. Il committer viene utilizzato per i processi Spark che usano Spark SQL, DataFrames o set di dati. A partire da Amazon EMR 6.4.0, questo committer può essere utilizzato per tutti i formati comuni, tra cui Parquet, ORC e formati testuali (inclusi CSV e JSON). Per i rilasci precedenti ad Amazon EMR 6.4.0, è supportato solo il formato Parquet. Vi sono circostanze in cui il committer non viene utilizzato. Per ulteriori informazioni, consulta [Requisiti per il committer ottimizzato S3 EMRFS](#).

Argomenti

- [Requisiti per il committer ottimizzato S3 EMRFS](#)

- [Committer ottimizzato S3 EMRFS e caricamenti in più parti](#)
- [Considerazioni sull'ottimizzazione dei processi](#)
- [Abilitare il committer ottimizzato S3 EMRFS per Amazon EMR 5.19.0](#)

Requisiti per il committer ottimizzato S3 EMRFS

Il committer ottimizzato S3 EMRFS viene utilizzato quando si verificano le condizioni riportate di seguito:

- Esegui i processi Spark che usano Spark SQL, DataFrames o set di dati per scrivere file ad Amazon S3. A partire da Amazon EMR 6.4.0, questo committer può essere utilizzato per tutti i formati comuni, tra cui Parquet, ORC e formati testuali (inclusi CSV e JSON). Per i rilasci precedenti ad Amazon EMR 6.4.0, è supportato solo il formato Parquet.
- I caricamenti in più parti sono abilitati in Amazon EMR. Questa è l'impostazione predefinita. Per ulteriori informazioni, consulta [Committer ottimizzato S3 EMRFS e caricamenti in più parti](#).
- Viene utilizzato il supporto integrato per i formati di file di Spark. Il supporto integrato per i formati viene utilizzato nelle seguenti circostanze:
 - Per le tabelle Hive metastore, quando `spark.sql.hive.convertMetastoreParquet` è impostato su `true` per le tabelle Parquet, oppure quando `spark.sql.hive.convertMetastoreOrc` è impostato su `true` per le tabelle Orc con EMR rilascio 6.4.0 o successivi. Queste sono le impostazioni predefinite.
 - Quando i processi scrivono in origini dei dati o tabelle di formato di file, ad esempio la tabella di destinazione viene creata con la clausola `USING parquet`.
 - Quando i processi scrivono tabelle Parquet non partizionate Hive metastore. Il supporto per Parquet integrato di Spark non supporta le tabelle Hive partizionate, che è una limitazione nota. Per ulteriori informazioni, consulta la sezione relativa alla [conversione delle tabelle Parquet del metastore Hive](#) nel manuale Apache Spark SQL, DataFrames and Datasets Guide.
- Le operazioni di processo Spark che scrivono in una posizione di partizione predefinita, ad esempio `/${table_location}/k1=v1/k2=v2/`, usano il committer. Il committer non viene utilizzato se un'operazione di processo scrive in una posizione di partizione personalizzata, ad esempio se un percorso di partizione personalizzato è impostato utilizzando il comando `ALTER TABLE SQL`.
- Devono essere utilizzati i seguenti valori per Spark:
 - La proprietà `spark.sql.parquet.fs.optimized.committer.optimization-enabled` deve essere impostata su `true`. Questa è l'impostazione predefinita con Amazon EMR 5.20.0 e

versione successiva. Con Amazon EMR 5.19.0, il valore predefinito è `false`. Per informazioni su come configurare questo valore, consulta [Abilitare il committer ottimizzato S3 EMRFS per Amazon EMR 5.19.0](#).

- Se si scrive su tabelle di metastore Hive non partizionate, sono supportati solo i formati di file Parquet e Orc. `spark.sql.hive.convertMetastoreParquet` deve essere impostato su `true` se la scrittura avviene su tabelle di metastore Parquet Hive non partizionate. `spark.sql.hive.convertMetastoreOrc` deve essere impostato su `true` se la scrittura avviene su tabelle di metastore Orc Hive non partizionate. Queste sono le impostazioni predefinite.
- `spark.sql.parquet.output.committer.class` deve essere impostato su `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter`. Questa è l'impostazione predefinita.
- `spark.sql.sources.commitProtocolClass` deve essere impostato su `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` o `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`. `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` è l'impostazione predefinita per la serie EMR 5.x versione 5.30.0 e successive e per la serie EMR 6.x versione 6.2.0 e successive. `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` è l'impostazione predefinita per le versioni EMR precedenti.
- Se i processi Spark sovrascrivono i set di dati Parquet partizionati con colonne di partizione dinamiche, le opzioni di scrittura `partitionOverwriteMode` e `spark.sql.sources.partitionOverwriteMode` devono essere impostate su `static`. Questa è l'impostazione predefinita.

Note

L'opzione di scrittura `partitionOverwriteMode` è stata introdotta in Spark 2.4.0. Per Spark versione 2.3.2, incluso con Amazon EMR rilascio 5.19.0, imposta la proprietà `spark.sql.sources.partitionOverwriteMode`.

Quando il committer ottimizzato S3 EMRFS non viene utilizzato

In genere, il committer EMRFS ottimizzato per S3 non viene utilizzato nelle situazioni riportate di seguito.

Situazione	Perché il committer non viene utilizzato
Quando si scrive su HDFS	Il committer supporta solo la scrittura su Amazon S3 utilizzando EMRFS.
Quando si utilizza il file system S3A	Il committer supporta solo EMRFS.
Quando si utilizza MapReduce o l'API RDD di Spark	Il committer supporta solo l'utilizzo di API SparkSQL, DataFrame o Dataset.

I seguenti esempi di Scala mostrano altre situazioni che impediscono di utilizzare il committer EMRFS ottimizzato per S3 in tutto (il primo esempio) e in parte (il secondo esempio).

Example - Modalità di sovrascrittura dinamica delle partizioni

Il seguente esempio di Scala indica a Spark di utilizzare un algoritmo di commit diverso, che impedisce l'uso del committer EMRFS ottimizzato per S3. Il codice imposta la proprietà `partitionOverwriteMode` su `dynamic` per sovrascrivere solo le partizioni su cui si stanno scrivendo i dati. Quindi, le colonne delle partizioni dinamiche vengono specificate da `partitionBy` e la modalità scrittura è impostata su `overwrite`.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")
  .option("partitionOverwriteMode", "dynamic")
  .partitionBy("dt")
  .parquet("s3://EXAMPLE-DOC-BUCKET/output")
```

È necessario configurare tutte e tre le impostazioni per evitare l'utilizzo del committer EMRFS ottimizzato per S3. In questo modo, Spark esegue un algoritmo di commit diverso specificato nel protocollo di commit di Spark. Per i rilasci di Amazon EMR serie 5.x precedenti al 5.30.0 e per i rilasci di Amazon EMR serie 6.x precedenti a 6.2.0, il protocollo di commit utilizza la directory di gestione temporanea di Spark, che è una directory temporanea creata nella posizione di output che inizia con

.spark-staging. L'algoritmo rinomina in modo sequenziale le directory delle partizioni e questo può influire negativamente sulle prestazioni. Per ulteriori informazioni su Amazon EMR rilascio 5.30.0 e successivi e sul rilascio 6.2.0 e successivi, consulta la sezione [Utilizzo del protocollo di commit ottimizzato per S3 EMRFS](#).

L'algoritmo in Spark 2.4.0 segue questi passaggi:

1. I tentativi di attività scrivono il loro output nelle directory delle partizioni sotto la directory di staging di Spark, ad esempio `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`.
2. Per ogni partizione scritta, il tentativo di attività tiene traccia dei percorsi di partizione relativi, ad esempio `k1=v1/k2=v2`.
3. Quando un'attività viene completata correttamente, fornisce al driver tutti i relativi percorsi di partizione che ha tracciato.
4. Al termine di tutte le attività, la fase di commit dei lavori raccoglie tutte le directory delle partizioni che i tentativi di attività riusciti hanno scritto nella directory di gestione temporanea di Spark. Spark rinomina in sequenza ciascuna di queste directory nella sua posizione di output finale utilizzando le operazioni di ridenominazione dell'albero delle directory.
5. La directory di gestione temporanea viene eliminata prima del completamento della fase di commit del processo.

Example - Posizione della partizione personalizzata

In questo esempio, il codice Scala viene inserito in due partizioni. Una partizione ha una posizione di partizione personalizzata. L'altra partizione usa il percorso di partizione predefinito. Il committer ottimizzato S3 EMRFS viene utilizzato solo per scrivere l'output dell'attività nella partizione che utilizza la posizione della partizione predefinita.

```
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)

```

Il codice Scala crea i seguenti oggetti Amazon S3:

```

custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_$folder$
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_$folder$
table_$folder$

```

Quando si scrive nelle partizioni in posizioni personalizzate, Spark utilizza un algoritmo di commit simile all'esempio precedente, che è descritto di seguito. Come nell'esempio precedente, l'algoritmo produce ridenominazioni sequenziali che possono influire negativamente sulle prestazioni.

1. Quando si scrive l'output in una partizione in una posizione personalizzata, le attività scrivono in un file nella directory di gestione temporanea di Spark, che viene creata nella posizione di output finale. Il nome del file include un UUID casuale per proteggere il file da collisioni. Il tentativo di attività tiene traccia di ogni file insieme al percorso di output finale desiderato.
2. Quando un'attività viene completata correttamente, fornisce al driver i file e i relativi percorsi di output finali desiderati.
3. Al termine di tutte le attività, la fase di commit dei lavori consente di rinominare in sequenza tutti i file scritti per le partizioni in percorsi personalizzati nei relativi percorsi di output finali.
4. La directory di gestione temporanea viene eliminata prima del completamento della fase di commit del processo.

Committer ottimizzato S3 EMRFS e caricamenti in più parti

Per utilizzare il committer ottimizzato S3 EMRFS, i caricamenti in più parti devono essere abilitati in Amazon EMR. I caricamenti in più parti sono abilitati per impostazione predefinita. È possibile abilitarli nuovamente, se necessario. Per ulteriori informazioni, consulta [Configurazione del caricamento in più parti per Amazon S3](#) nella Guida alla gestione di Amazon EMR.

Il committer ottimizzato S3 EMRFS utilizza le caratteristiche tipo transazione dei caricamenti in più parti per garantire che i file scritti da tentativi di attività siano visualizzati solo nella posizione di output del processo durante il commit delle attività. Utilizzando i caricamenti in più parti in questo modo, il committer migliora le prestazioni delle attività di commit sull'algoritmo `FileOutputCommitter` versione 2. Quando utilizzi il committer ottimizzato S3 EMRFS, ci sono delle differenze chiave rispetto al comportamento tradizionale del caricamento in più parti di cui tenere conto:

- I caricamenti in più parti vengono sempre eseguiti indipendentemente dalle dimensioni del file. Questo differisce dal comportamento predefinito di EMRFS, dove la proprietà `fs.s3n.multipart.uploads.split.size` controlla le dimensioni del file in cui vengono attivati i caricamenti in più parti.
- I caricamenti in più parti restano in stato incompleto per un periodo di tempo più lungo fino a quando avviene il commit o l'arresto dell'attività. Questo differisce dal comportamento predefinito di EMRFS dove un caricamento in più parti si completa quando l'attività termina la scrittura in un determinato file.

A causa di queste differenze, se un Executor JVM Spark si arresta in modo anomalo o viene interrotto mentre le attività sono in corso e stanno scrivendo dati in Amazon S3, i caricamenti in più parti hanno più probabilità di restare indietro. Per questo motivo, quando utilizzi il committer ottimizzato S3 EMRFS, assicurati di seguire le best practice per la gestione dei caricamenti in più parti non riusciti. Per ulteriori informazioni, consulta le [best practice](#) per utilizzare i bucket Amazon S3 nella Guida alla gestione di Amazon EMR.

Considerazioni sull'ottimizzazione dei processi

Il committer ottimizzato S3 EMRFS consuma una piccola quantità di memoria per ogni file scritto da un tentativo di attività fino a quando avviene il commit o l'arresto dell'attività. Nella maggior parte dei processi la quantità di memoria consumata è trascurabile. Per i processi che presentano attività di lunga esecuzione che scrivono un numero elevato di file, la memoria che il committer consuma può essere notevole e richiede l'adeguamento della memoria allocata per gli executor Spark. Puoi ottimizzare la memoria dell'executor mediante la proprietà `spark.executor.memory`.

Orientativamente, un'unica attività che scrive 100.000 file in genere richiede ulteriori 100 MB di memoria. Per ulteriori informazioni, consulta le [Proprietà delle applicazioni](#) nella documentazione relativa alla configurazione di Apache Spark.

Abilitare il committer ottimizzato S3 EMRFS per Amazon EMR 5.19.0

Se usi Amazon EMR 5.19.0, puoi impostare manualmente la proprietà `spark.sql.parquet.fs.optimized.committer.optimization-enabled` su `true` quando crei un cluster o da Spark se stai utilizzando Amazon EMR.

Abilitare il committer ottimizzato S3 EMRFS al momento della creazione di un cluster

È possibile utilizzare la classificazione di configurazione `spark-defaults` per impostare la proprietà `spark.sql.parquet.fs.optimized.committer.optimization-enabled` su `true`. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Abilitare il committer ottimizzato S3 EMRFS da Spark

È possibile impostare `spark.sql.parquet.fs.optimized.committer.optimization-enabled` su `true` impostandola come `hardcoded` in un `SparkConf`, passandola come parametro `--conf` nella shell Spark o come strumenti `spark-submit` e `spark-sql`, o in `conf/spark-defaults.conf`. Per ulteriori informazioni, consulta la sezione relativa alla [Spark configuration \(Configurazione Spark\)](#) nella documentazione di Apache Spark.

L'esempio seguente mostra come abilitare il committer durante l'esecuzione di un comando `spark-sql`.

```
spark-sql \  
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \  
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

Utilizzo del protocollo di commit ottimizzato per S3 EMRFS

Il protocollo di commit ottimizzato per S3 EMRFS è un'alternativa all'implementazione di [FileCommitProtocol](#) che è ottimizzata per scrivere i file con la sovrascrittura dinamica delle partizioni di Spark in Amazon S3 quando si usa EMRFS. Il protocollo migliora le prestazioni delle applicazioni evitando le operazioni di ridenominazione in Amazon S3 durante la fase di commit del processo di sovrascrittura dinamica delle partizioni di Spark.

Tieni presente che la [Utilizzare il committer ottimizzato S3 EMRFS](#) migliora le prestazioni anche evitando le operazioni di ridenominazione. Tuttavia, non funziona per i casi di sovrascrittura dinamica

delle partizioni, mentre i miglioramenti del protocollo commit riguardano solo i casi di sovrascrittura dinamica delle partizioni.

Il protocollo di commit è disponibile in Amazon EMR rilascio 5.30.0 e successivi ed è abilitato per impostazione predefinita in Amazon EMR rilascio 6.2.0 e successivi. Amazon EMR ha aggiunto un miglioramento del parallelismo a partire dal rilascio 5.31.0. Il protocollo viene utilizzato per i processi Spark che impiegano Spark SQL, DataFrames o set di dati. Esistono circostanze in cui il protocollo di commit non viene utilizzato. Per ulteriori informazioni, consulta [Requisiti per il protocollo di commit ottimizzato per S3 EMRFS](#).

Argomenti


- [Requisiti per il protocollo di commit ottimizzato per S3 EMRFS](#)
- [Protocollo di commit ottimizzato per S3 EMRFS e caricamenti in più parti](#)
- [Considerazioni sull'ottimizzazione dei processi](#)

Requisiti per il protocollo di commit ottimizzato per S3 EMRFS

Il protocollo di commit ottimizzato per S3 EMRFS viene utilizzato quando si verificano le condizioni riportate di seguito:

- Vengono eseguiti processi Spark che impiegano Spark SQL, DataFrames o set di dati per sovrascrivere tabelle partizionate.
- Vengono eseguiti processi Spark la cui modalità di sovrascrittura delle partizioni è `dynamic`.
- I caricamenti in più parti sono abilitati in Amazon EMR. Questa è l'impostazione predefinita. Per ulteriori informazioni, consulta [Protocollo di commit ottimizzato per S3 EMRFS e caricamenti in più parti](#).
- La cache del file system per EMRFS è abilitata. Questa è l'impostazione predefinita. Verifica che l'impostazione `fs.s3.impl.disable.cache` sia impostata su `false`.
- Viene utilizzato il supporto integrato per le origini dei dati di Spark. Il supporto integrato per le origini dei dati viene utilizzato nelle seguenti circostanze:
 - Quando i processi scrivono su origini dei dati o tabelle integrate.
 - Quando i processi scrivono tabelle Parquet del metastore Hive. Ciò accade quando `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastoreParquet` sono entrambi impostati su `true`. Queste sono le impostazioni predefinite.

- Quando i processi scrivono tabelle ORC del metastore Hive. Ciò accade quando `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastoreOrc` sono entrambi impostati su `true`. Queste sono le impostazioni predefinite.
- Le operazioni di processo Spark che scrivono in una posizione di partizione predefinita, ad esempio `/${table_location}/k1=v1/k2=v2/`, usano il protocollo di commit. Il protocollo non viene utilizzato se un'operazione di processo scrive in una posizione di partizione personalizzata, ad esempio se un percorso di partizione personalizzato è impostato utilizzando il comando `ALTER TABLE SQL`.
- Devono essere utilizzati i seguenti valori per Spark:
 - `spark.sql.sources.commitProtocolClass` deve essere impostato su `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`. Questa è l'impostazione predefinita per Amazon EMR rilascio 5.30.0 e successivi, e rilascio 6.2.0 e successivi.
 - L'opzione di scrittura `partitionOverwriteMode` o `spark.sql.sources.partitionOverwriteMode` deve essere impostata su `dynamic`. L'impostazione predefinita è `static`.

 Note

L'opzione di scrittura `partitionOverwriteMode` è stata introdotta in Spark 2.4.0. Per Spark versione 2.3.2, incluso con Amazon EMR rilascio 5.19.0, imposta la proprietà `spark.sql.sources.partitionOverwriteMode`.

- Se i processi Spark sovrascrivono la tabella Parquet del metastore Hive, `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastore.partitionOverwriteMode` devono essere impostati su `true`. Queste sono le impostazioni predefinite.
- Se i processi Spark sovrascrivono la tabella ORC del metastore Hive, `spark.sql.hive.convertMetastoreOrc`, `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastore.partitionOverwriteMode` devono essere impostati su `true`. Queste sono le impostazioni predefinite.

Example - Modalità di sovrascrittura dinamica delle partizioni

In questo esempio di Scala, viene attivata l'ottimizzazione. Innanzitutto, imposta la proprietà `partitionOverwriteMode` su `dynamic`. Questo sovrascrive solo le partizioni in cui stai scrivendo i dati. Quindi, specifichi le colonne delle partizioni dinamiche con `partitionBy` e imposti la modalità scrittura su `overwrite`.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")           // "overwrite" instead of "insert"
  .option("partitionOverwriteMode", "dynamic") // "dynamic" instead of "static"
  .partitionBy("dt")                       // partitioned data instead of
  unpartitioned data
  .parquet("s3://EXAMPLE-DOC-BUCKET/output") // "s3://" to use EMR file system,
  instead of "s3a://" or "hdfs://"
```

Quando il protocollo di commit ottimizzato per S3 EMRFS non viene utilizzato

In genere, il protocollo di commit ottimizzato per S3 EMRFS funziona allo stesso modo del protocollo di commit SQL Spark predefinito open source, `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`. L'ottimizzazione non si verifica nelle seguenti situazioni.

Situazione	Perché il protocollo di commit non viene utilizzato
Quando si scrive su HDFS	Il protocollo di commit supporta solo la scrittura su Amazon S3 utilizzando EMRFS.
Quando si utilizza il file system S3A	Il protocollo di commit supporta solo EMRFS.
Quando si utilizza MapReduce o l'API RDD di Spark	Il protocollo di commit supporta solo l'utilizzo delle API di SparkSQL, DataFrame o set di dati.
Quando la sovrascrittura dinamica delle partizioni non viene attivata	Il protocollo di commit ottimizza solo i casi di sovrascrittura dinamica delle partizioni. Per altri casi, consulta la sezione Utilizzare il committer ottimizzato S3 EMRFS .

I seguenti esempi di Scala mostrano altre situazioni che il protocollo di commit ottimizzato per S3 EMRFS delega `SQLHadoopMapReduceCommitProtocol`.

Example - Modalità di sovrascrittura delle partizioni con posizione della partizione personalizzata

In questo esempio, i programmi Scala sovrascrivono due partizioni in modalità di sovrascrittura dinamica delle partizioni. Una partizione ha una posizione di partizione personalizzata. L'altra partizione usa il percorso di partizione predefinito. Il protocollo di commit ottimizzato per S3 EMRFS viene utilizzato solo per scrivere l'output dell'attività nella partizione che utilizza la posizione della partizione predefinita.

```
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")
```

Il codice Scala crea i seguenti oggetti Amazon S3:

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder$}
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder$}
table_${folder$}
```

Note

La scrittura in posizioni di partizione personalizzate nelle versioni precedenti di Spark può causare la perdita di dati. In questo esempio, la partizione `dt= '2019-01-28'` andrebbe persa. Per ulteriori dettagli, consulta [SPARK-35106](#). Questo problema è stato risolto in Amazon EMR rilascio 5.33.0 e successivi, esclusi i rilasci 6.0.x e 6.1.x.

Quando si scrive nelle partizioni in posizioni personalizzate, Spark utilizza un algoritmo di commit simile all'esempio precedente, che è descritto di seguito. Come nell'esempio precedente, l'algoritmo produce ridenominazioni sequenziali che possono influire negativamente sulle prestazioni.

L'algoritmo in Spark 2.4.0 segue questi passaggi:

1. Quando si scrive l'output in una partizione in una posizione personalizzata, le attività scrivono in un file nella directory di gestione temporanea di Spark, che viene creata nella posizione di output finale. Il nome del file include un UUID casuale per proteggere il file da collisioni. Il tentativo di attività tiene traccia di ogni file insieme al percorso di output finale desiderato.
2. Quando un'attività viene completata correttamente, fornisce al driver i file e i relativi percorsi di output finali desiderati.
3. Al termine di tutte le attività, la fase di commit dei lavori consente di rinominare in sequenza tutti i file scritti per le partizioni in percorsi personalizzati nei relativi percorsi di output finali.
4. La directory di gestione temporanea viene eliminata prima del completamento della fase di commit del processo.

Protocollo di commit ottimizzato per S3 EMRFS e caricamenti in più parti

Per utilizzare l'ottimizzazione per la sovrascrittura dinamica delle partizioni nel protocollo di commit ottimizzato per S3 EMRFS, i caricamenti in più parti devono essere abilitati in Amazon EMR. I

caricamenti in più parti sono abilitati per impostazione predefinita. È possibile abilitarli nuovamente, se necessario. Per ulteriori informazioni, consulta [Configurazione del caricamento in più parti per Amazon S3](#) nella Guida alla gestione di Amazon EMR.

Durante la scovrascrittura dinamica delle partizioni, il protocollo di commit ottimizzato per S3 EMRFS utilizza le caratteristiche tipo transazione dei caricamenti in più parti per garantire che i file scritti da tentativi di attività siano visualizzati solo nella posizione di output del processo durante il commit del processo. Utilizzando i caricamenti in più parti in questo modo, il protocollo di commit migliora le prestazioni del commit dei processi rispetto al protocollo `SQLHadoopMapReduceCommitProtocol` predefinito. Quando utilizzi il protocollo di commit ottimizzato per S3 EMRFS, esistono delle differenze chiave rispetto al comportamento tradizionale del caricamento in più parti di cui tenere conto:

- I caricamenti in più parti vengono sempre eseguiti indipendentemente dalle dimensioni del file. Questo differisce dal comportamento predefinito di EMRFS, dove la proprietà `fs.s3n.multipart.uploads.split.size` controlla le dimensioni del file in cui vengono attivati i caricamenti in più parti.
- I caricamenti in più parti restano in stato incompleto per un periodo di tempo più lungo fino a quando avviene il commit o l'arresto dell'attività. Questo differisce dal comportamento predefinito di EMRFS dove un caricamento in più parti si completa quando l'attività termina la scrittura in un determinato file.

A causa di queste differenze, se un esecutore JVM di Spark si arresta in modo anomalo o viene interrotto mentre le attività sono in corso e stanno scrivendo dati in Amazon S3, oppure un driver JVM di Spark ha un arresto anomalo o viene chiuso in modo forzato mentre è in corso un processo, i caricamenti in più parti hanno più probabilità di restare indietro. Per questo motivo, quando utilizzi il protocollo di commit ottimizzato per S3 EMRFS, assicurati di seguire le best practice per la gestione dei caricamenti in più parti non riusciti. Per ulteriori informazioni, consulta le [best practice](#) per utilizzare i bucket Amazon S3 nella Guida alla gestione di Amazon EMR.

Considerazioni sull'ottimizzazione dei processi

Sugli esecutori Spark, il protocollo di commit ottimizzato per S3 EMRFS consuma una piccola quantità di memoria per ogni file scritto da un tentativo di attività fino a quando avviene il commit o l'arresto dell'attività. Nella maggior parte dei processi la quantità di memoria consumata è trascurabile.

Sui driver Spark, il protocollo di commit ottimizzato per S3 EMRFS richiede memoria per archiviare le informazioni sui metadati di ogni file salvato fino a quando avviene il commit o l'arresto del processo. Nella maggior parte dei processi, l'impostazione predefinita della memoria del driver Spark è trascurabile.

Per i processi che presentano attività di lunga esecuzione che scrivono un numero elevato di file, la memoria che il protocollo di commit consuma può essere notevole e rende necessario adeguare la memoria allocata per Spark, specialmente per gli esecutori Spark. È possibile regolare la memoria utilizzando la proprietà `spark.driver.memory` per i driver Spark e la proprietà `spark.executor.memory` per gli esecutori Spark. Orientativamente, un'unica attività che scrive 100.000 file in genere richiede ulteriori 100 MB di memoria. Per ulteriori informazioni, consulta le [Proprietà delle applicazioni](#) nella documentazione relativa alla configurazione di Apache Spark.

Effettuare nuovi tentativi per le richieste Amazon S3 con EMRFS

In questo argomento vengono fornite informazioni sulle strategie di nuovi tentativi che è possibile utilizzare quando si effettuano richieste ad Amazon S3 con EMRFS. Quando il tasso di richiesta aumenta, S3 tenta di dimensionarsi per supportare la nuova tariffa. Durante questo processo, S3 può limitare le richieste e restituire un errore `503 Slow Down`. Per migliorare il tasso di successo delle richieste S3, è possibile modificare la strategia di nuovi tentativi impostando le proprietà nella configurazione di `emrfs-site`

È possibile modificare la strategia di nuovi tentativi nei modi seguenti.

- Aumentare il limite massimo di tentativi per la strategia di nuovi tentativi con backoff esponenziale predefinita.
- Abilitare e configurare la strategia di nuovi tentativi come aumento additivo - diminuzione moltiplicativa (AIMD). AIMD è supportato per Amazon EMR rilascio 6.4.0 e successivi.

Usare la strategia di backoff esponenziale predefinita

Per impostazione predefinita, EMRFS utilizza una strategia di backoff esponenziale per effettuare nuovi tentativi per le richieste Amazon S3. Il limite di nuovi tentativi predefinito di EMRFS è pari a 15. Per evitare un errore `503 Slow Down` di S3 è possibile aumentare il limite di nuovi tentativi quando si crea un nuovo cluster, in un cluster in esecuzione o in fase di runtime dell'applicazione.

Per aumentare il limite di nuovi tentativi è necessario modificare il valore di `fs.s3.maxRetries` nella propria configurazione `emrfs-site`. I seguenti set di configurazione di esempio impostano `fs.s3.maxRetries` a un valore personalizzato pari a 30.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.maxRetries": "30"
    }
  }
]
```

Per ulteriori informazioni sull'utilizzo degli oggetti di configurazione, consulta [Configurazione delle applicazioni](#).

Usare la strategia di nuovi tentativi dell'AIMD

Con Amazon EMR rilascio 6.4.0 e successivi, EMRFS supporta una strategia alternativa di ripetizione basata su un modello di aumento additivo/diminuzione moltiplicativa (AIMD). La strategia di nuovi tentativi AIMD è particolarmente utile quando si lavora con cluster Amazon EMR di grandi dimensioni.

AIMD calcola una frequenza di richiesta personalizzata utilizzando i dati sulle recenti richieste di successo. Questa strategia riduce il numero di richieste limitate e i tentativi totali richiesti per ciascuna richiesta.

Per abilitare la strategia di nuovi tentativi AIMD, è necessario impostare la proprietà `fs.s3.aimd.enabled` su `true` nella configurazione `emrfs-site` come nell'esempio seguente.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.aimd.enabled": "true"
    }
  }
]
```

Per ulteriori informazioni sull'utilizzo degli oggetti di configurazione, consulta [Configurazione delle applicazioni](#).

Impostazioni avanzate di nuovi tentativi dell'AIMD

È possibile configurare le proprietà elencate nella tabella seguente per perfezionare il comportamento dei nuovi tentativi quando si utilizza la strategia di nuovi tentativi dell'AIMD. Per la maggior parte dei casi d'uso, si consiglia di utilizzare i valori predefiniti.

Proprietà avanzate della strategia di nuovi tentativi dell'AIMD

Proprietà	Valore predefinito	Descrizione
<code>fs.s3.aimd.increaseIncrement</code>	0.1	Controlla la rapidità con cui il tasso di richiesta aumenta quando le richieste consecutive hanno esito positivo.
<code>fs.s3.aimd.reductionFactor</code>	2	Controlla la rapidità con cui il tasso di richiesta diminuisce quando Amazon S3 restituisce una risposta 503. Il fattore predefinito di 2 riduce a metà il tasso di richiesta.
<code>fs.s3.aimd.minRate</code>	0.1	Imposta il limite inferiore per il tasso di richiesta quando le richieste subiscono una limitazione sostenuta da S3.
<code>fs.s3.aimd.initialRate</code>	5500	Imposta il tasso di richiesta iniziale, che cambia in base ai valori specificati per <code>fs.s3.aimd.increaseIncrement</code> e <code>fs.s3.aimd.reductionFactor</code> . La frequenza iniziale viene utilizzata anche per le richieste GET e ridimensionata

Proprietà	Valore predefinito	Descrizione
		proporzionalmente (3500/5500) per le richieste PUT.
<code>fs.s3.aimd.adjustWindow</code>	2	Controlla la frequenza di adeguamento del tasso di richiesta, misurato in numero di risposte.
<code>fs.s3.aimd.maxAttempts</code>	100	Imposta il numero massimo di tentativi di provare una richiesta.

Aggiungi una fase Spark

Puoi utilizzare le fasi di Amazon EMR per inviare lavoro al framework Spark installato su un cluster EMR. Per ulteriori informazioni, consulta l'argomento [Fasi](#) della Guida alla gestione di Amazon EMR. Nella console e nella CLI, puoi eseguire questa operazione utilizzando una fase dell'applicazione Spark, che esegue automaticamente lo script `spark-submit` come fase. Con l'API, utilizza una fase per invocare `spark-submit` utilizzando `command-runner.jar`.

Per ulteriori informazioni su come inviare applicazioni a Spark, consulta l'argomento [Submitting applications \(Invio di applicazioni\)](#) nella documentazione di Apache Spark.

Per inviare una fase Spark mediante la console

1. Apri la console di Amazon EMR all'indirizzo <https://console.aws.amazon.com/emr>.
2. In Cluster List (Elenco cluster), scegliere il nome del cluster.
3. Scorrere fino alla sezione Steps (Fasi) ed espanderla, quindi scegliere Add step (Aggiungi fase).
4. Nella finestra di dialogo Add Step (Aggiungi fase):
 - Per Tipo di fase, scegli Applicazione Spark.
 - Per Name (Nome), accettare il nome predefinito (applicazione Spark) o digitare un nuovo nome.
 - Per Deploy mode (Modalità di distribuzione), scegliere la modalità Client o Cluster. La modalità Client avvia il programma driver nell'istanza primaria del cluster, mentre la modalità Cluster

avvia il programma driver nel cluster. Per la modalità client, l'output del log del driver viene visualizzato nei log delle fasi, mentre per la modalità cluster, l'output del log del driver viene visualizzato nei log per il primo container YARN. Per ulteriori informazioni, consulta [Cluster mode overview \(Panoramica della modalità cluster\)](#) nella documentazione di Apache Spark.

- Specificare le Spark-submit options (Opzioni Spark-submit) desiderate. Per ulteriori informazioni sulle opzioni `spark-submit`, consulta [Avvio di applicazioni con spark-submit](#).
 - Per Application location (Posizione applicazione), specificare il percorso locale o URI di S3 dell'applicazione.
 - Lasciare vuoto il campo Arguments (Argomenti).
 - Per Action on failure (Operazione in caso di errore), accettare l'opzione predefinita, ovvero Continue (Continua).
5. Scegliere Add (Aggiungi). La fase viene visualizzata nella console con lo stato Pending (In attesa).
 6. Durante l'esecuzione della fase, lo stato passa da Pending (In attesa) a Running (In esecuzione) a Completed (Completata). Per aggiornare lo stato, scegliere l'icona Refresh (Aggiorna) sopra la colonna Actions (Operazioni).
 7. I risultati della fase sono visualizzati nella pagina Cluster Details (Dettagli cluster) della console di Amazon EMR accanto alla fase in Log Files (File di log) se la caratteristica di registrazione dei log è stata configurata. È eventualmente possibile trovare informazioni sulla fase nel bucket di log configurato all'avvio del cluster.

Per inviare lavoro a Spark mediante il kit AWS CLI

Inviare una fase quando si crea il cluster o si utilizza il sottocomando `aws emr add-steps` in un cluster esistente.

1. Utilizzare `create-cluster` come mostrato nell'esempio seguente.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-5.36.1
  --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
```

È anche possibile utilizzare `command-runner.jar` come mostrato nell'esempio seguente.

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-5.36.1 \
  --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

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

2. In alternativa, aggiungere fasi a un cluster in esecuzione. Utilizza `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]
```

È anche possibile utilizzare `command-runner.jar` come mostrato nell'esempio seguente.

```
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]
```

Invio di lavoro a Spark mediante il kit SDK per Java

1. L'esempio seguente mostra come aggiungere una fase a un cluster con Spark utilizzando Java.

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

StepFactory stepFactory = new StepFactory();
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig sparkStepConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("spark-submit", "--executor-memory", "1g", "--
class", "org.apache.spark.examples.SparkPi", "/usr/lib/spark/examples/jars/spark-
examples.jar", "10");

StepConfig sparkStep = new StepConfig()
    .withName("Spark Step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(sparkStepConf);

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

2. Visualizzare i risultati della fase esaminando i log relativi alla stessa. È possibile eseguire questa operazione nella AWS Management Console se è stata attivata la registrazione scegliendo Steps (Fasi), selezionando la fase e quindi scegliendo o per stdoutLog Files (File di log)stderr. Per visualizzare i log disponibili, scegliere View Logs (Visualizza log).

Sostituzione delle impostazioni di configurazione predefinite di Spark

I valori di configurazione predefiniti di Spark possono essere sostituiti per ogni applicazione. Puoi eseguire questa operazione quando invii applicazioni utilizzando una fase, che in pratica passa opzioni a `spark-submit`. Ad esempio, puoi cambiare la memoria assegnata a un processo esecutore modificando `spark.executor.memory`. In tal caso, devi fornire il parametro `--executor-memory` con un argomento come il seguente:

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/  
spark/examples/jars/spark-examples.jar 10
```

Nello stesso modo, puoi regolare `--executor-cores` e `--driver-memory`. In una fase, forniresti i seguenti argomenti alla fase:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/  
jars/spark-examples.jar 10
```

Puoi anche regolare impostazioni che possono non avere un parametro integrato utilizzando l'opzione `--conf`. Per ulteriori informazioni su altre impostazioni regolabili, consulta l'argomento [Dynamically loading Spark properties \(Caricamento dinamico delle proprietà Spark\)](#) nella documentazione di Apache Spark.

Visualizzazione della cronologia delle applicazioni Spark

È possibile visualizzare i dettagli dell'interfaccia utente di Spark, dell'applicazione YARN e dell'interfaccia utente di Tez utilizzando la scheda Application user interfaces (Interfacce utente delle applicazioni) nella pagina dei dettagli di un cluster nella console. Le interfacce utente (UI) dell'applicazione Amazon EMR facilitano la risoluzione dei problemi e l'analisi dei processi attivi e della cronologia dei processi.

Per ulteriori informazioni, consulta [Visualizzazione della cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.

Accedi alle interfacce utente Web di Spark

È possibile visualizzare le interfacce utente Web di Spark seguendo le procedure per creare un tunnel SSH o un proxy incluse nella sezione [Connessione al cluster](#) nella Guida alla gestione di Amazon EMR e quindi accedendo al servizio YARN ResourceManager per il cluster. Scegliere il collegamento per l'applicazione in Tracking UI (Monitoraggio IU). Se l'applicazione è in esecuzione, viene visualizzato ApplicationMaster. Ciò consente di accedere all'interfaccia utente Web dell'applicazione master sulla porta 20888 indipendentemente dalla posizione del driver. Il driver può trovarsi nel nodo primario del cluster se l'esecuzione avviene in modalità client YARN. Se un'applicazione è eseguita in modalità cluster YARN, il driver si trova in ApplicationMaster per l'applicazione nel cluster. Se l'applicazione non è in esecuzione, viene visualizzato History (Cronologia), che consente di accedere all'interfaccia utente Spark HistoryServer sulla porta 18080

del nodo primario del cluster EMR. Ciò vale per le applicazioni già completate. È anche possibile accedere direttamente all'interfaccia utente Spark HistoryServer all'indirizzo <http://master-public-dns-name:18080/>.

Con Amazon EMR rilascio 5.25.0 o successivi, puoi accedere all'interfaccia utente del server della cronologia Spark dalla console senza configurare un proxy Web tramite una connessione SSH. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce utente dell'applicazione persistente](#).

Utilizzo dell'integrazione di Amazon Redshift per Apache Spark su Amazon EMR

Con Amazon EMR rilascio 6.4.0 e successivi, ogni immagine del rilascio include un connettore tra [Apache Spark](#) e Amazon Redshift. Con questo connettore, è possibile utilizzare facilmente Spark su Amazon EMR per elaborare i dati archiviati in Amazon Redshift. Per i rilasci di Amazon EMR da 6.4.0 a 6.8.0, l'integrazione si basa sul [connettore spark-redshift open source](#). Per Amazon EMR rilascio 6.9.0 e successivi, l'[integrazione di Amazon Redshift per Apache Spark](#) è passata dalla versione della community a un'integrazione nativa.

Argomenti

- [Avvio di un'applicazione Spark utilizzando l'integrazione di Amazon Redshift per Apache Spark](#)
- [Autenticazione con l'integrazione di Amazon Redshift per Apache Spark](#)
- [Lettura e scrittura da e su Amazon Redshift](#)
- [Considerazioni e limitazioni relative all'utilizzo del connettore Spark](#)

Avvio di un'applicazione Spark utilizzando l'integrazione di Amazon Redshift per Apache Spark

A partire dalla versione 6.4 alla 6.9 di Amazon EMR, è necessario utilizzare l'opzione `--jars` o `--packages` per specificare quale dei seguenti file JAR si desidera utilizzare. L'opzione `--jars` specifica le dipendenze memorizzate localmente, in HDFS o utilizzando HTTP/S. Per visualizzare altre posizioni dei file supportate dall'opzione `--jars`, consulta la sezione [Advanced Dependency Management](#) (Gestione avanzata delle dipendenze) nella documentazione di Spark. L'opzione `--packages` specifica le dipendenze memorizzate nel repository pubblico Maven.

- `spark-redshift.jar`
- `spark-avro.jar`

- `RedshiftJDBC.jar`
- `minimal-json.jar`

Le versioni 6.10.0 e successive di Amazon EMR non richiedono la dipendenza `minimal-json.jar` e installano automaticamente le altre dipendenze su ciascun cluster per impostazione predefinita. Gli esempi seguenti mostrano come avviare un'applicazione Spark con l'integrazione di Amazon Redshift per Apache Spark.

Amazon EMR 6.10.0 +

L'esempio seguente mostra come avviare un'applicazione Spark con il connettore `spark-redshift` con le versioni 6.10 e successive di Amazon EMR.

```
spark-submit my_script.py
```

Amazon EMR 6.4.0 - 6.9.x

Per avviare un'applicazione Spark con un connettore `spark-redshift` dalla versione 6.4 alla 6.9 di Amazon EMR è necessario utilizzare l'opzione `--jars` o `--packages`, come mostrato nell'esempio seguente. Come vedrai, i percorsi elencati con l'opzione `--jars` sono i percorsi predefiniti per i file JAR.

```
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
```

Autenticazione con l'integrazione di Amazon Redshift per Apache Spark

Utilizzo di AWS Secrets Manager per recuperare le credenziali ed effettuare la connessione ad Amazon Redshift

L'esempio di codice riportato di seguito mostra come utilizzare AWS Secrets Manager per recuperare le credenziali per connettersi a un cluster Amazon Redshift con l'interfaccia PySpark per Apache Spark in Python.

```
from pyspark.sql import SQLContext
```

```
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

secretsmanager_client = boto3.client('secretsmanager')
secret_manager_response = secretsmanager_client.get_secret_value(
    SecretId='string',
    VersionId='string',
    VersionStage='string'
)
username = # get username from secret_manager_response
password = # get password from secret_manager_response
url = "jdbc:redshift://redshifthost:5439/database?user=" + username + "&password=" +
    password

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3://path/for/temp/data") \
    .load()
```

Utilizzo di IAM per recuperare le credenziali ed effettuare la connessione ad Amazon Redshift

È possibile utilizzare il driver JDBC versione 2 fornito da Amazon Redshift per connettersi all'endpoint Amazon Redshift con il connettore Spark. Per utilizzare AWS Identity and Access Management (IAM), configura il tuo URL JDBC per l'utilizzo dell'autenticazione IAM. Per connetterti a un cluster Redshift da Amazon EMR, devi autorizzare il tuo ruolo IAM a recuperare credenziali IAM temporanee. Assegna le seguenti autorizzazioni al tuo ruolo IAM in modo che possa recuperare le credenziali ed eseguire operazioni Amazon S3.

- [Redshift:GetClusterCredentials](#) (per cluster Redshift con provisioning)
- [Redshift:DescribeClusters](#) (per cluster Redshift con provisioning)
- [Redshift:GetWorkgroup](#) (per gruppi di lavoro Amazon Redshift Serverless)
- [Redshift:GetCredentials](#) (per gruppi di lavoro Amazon Redshift Serverless)
- [s3:GetBucket](#)
- [s3:GetBucketLocation](#)

- [s3:GetObject](#)
- [s3:PutObject](#)
- [s3:GetBucketLifecycleConfiguration](#)

Per ulteriori informazioni su `GetClusterCredentials`, consulta [Policy delle risorse per `GetClusterCredentials`](#).

Inoltre, devi assicurarti che Amazon Redshift possa assumere il ruolo IAM durante le operazioni COPY e UNLOAD.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "redshift.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

L'esempio seguente utilizza l'autenticazione IAM tra Spark e 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()
```

Lettura e scrittura da e su Amazon Redshift

I seguenti esempi di codice utilizzano PySpark per leggere e scrivere dati di esempio da e su un database Amazon Redshift con un'API di origine dei dati e tramite SparkSQL.

Data source API

Utilizza PySpark per leggere e scrivere dati di esempio da e su un database Amazon Redshift con un'API di origine dei dati.

```
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

Utilizza PySpark per leggere e scrivere dati di esempio da e su un database Amazon Redshift con SparkSQL.

```

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()

```

Considerazioni e limitazioni relative all'utilizzo del connettore Spark

- Si consiglia di attivare SSL per la connessione JDBC da Spark su Amazon EMR ad Amazon Redshift.
- Come best practice, è consigliabile gestire le credenziali per il cluster Amazon Redshift in AWS Secrets Manager. Per un esempio, consulta la sezione [Utilizzo di AWS Secrets Manager per recuperare le credenziali per la connessione ad Amazon Redshift](#).

- Si consiglia di passare un ruolo IAM con il parametro `aws_iam_role` per il parametro di autenticazione di Amazon Redshift.
- Il parametro `tempformat` attualmente non supporta il formato Parquet.
- L'URI `tempdir` indica una posizione Amazon S3. Questa directory temporanea non viene pulita in automatico e quindi potrebbe generare costi aggiuntivi.
- Prendi in considerazione i seguenti consigli per Amazon Redshift:
 - Si consiglia di bloccare l'accesso pubblico al cluster Amazon Redshift.
 - Si consiglia di attivare la [registrazione di log di verifica di Amazon Redshift](#).
 - Si consiglia di attivare la [crittografia dei dati inattivi di Amazon Redshift](#).
- Prendi in considerazione i seguenti consigli per Amazon S3:
 - Si consiglia di [bloccare l'accesso pubblico ai bucket Amazon S3](#).
 - Si consiglia di utilizzare la [crittografia lato server di Amazon S3](#) per crittografare i bucket Amazon S3 utilizzati.
 - Si consiglia di utilizzare le [policy del ciclo di vita di Amazon S3](#) per definire le regole di conservazione del bucket Amazon S3.
 - Amazon EMR verifica sempre il codice importato dall'open source nell'immagine. Per motivi di sicurezza, non supportiamo i seguenti metodi di autenticazione da Spark ad Amazon S3:
 - Impostazione di chiavi di accesso AWS nella classificazione di configurazione `hadoop-env`
 - Codifica di chiavi di accesso AWS nella URI `tempdir`

Per ulteriori informazioni sull'utilizzo del connettore e dei parametri supportati, consulta le seguenti risorse:

- [Amazon Redshift integration for Apache Spark](#) (Integrazione di Amazon Redshift per Apache Spark) nella Guida alla gestione di Amazon Redshift
- Il [repository della community spark-redshift](#) su Github

Cronologia dei rilasci di Spark

Nella seguente tabella sono elencati la versione di Spark inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

⚠ Important

Apache Spark versione 2.3.1, disponibile a partire da Amazon EMR rilascio 5.16.0, risolve [CVE-2018-8024](#) e [CVE-2018-1334](#). Si consiglia di eseguire la migrazione delle versioni precedenti di Spark a Spark 2.3.1 o versioni successive.

Informazioni sulla versione di Spark

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
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-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-6.13.0	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-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
		rn-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-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.11.1	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.11.0	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.10.1	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.10.0	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.9.1	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-6.9.0	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-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.8.1	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.8.0	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.7.0	3.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.1	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.36.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.6.0	3.2.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.35.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.5.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.4.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.3.1	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.3.0	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.2.1	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.2.0	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.1.1	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.1.0	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.0.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-6.0.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.34.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.33.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.33.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.32.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.32.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.31.1	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.31.0	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.30.2	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.30.1	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.30.0	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.29.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.28.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.28.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.27.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.27.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.26.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.25.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.24.1	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.24.0	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.23.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.23.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.22.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.2	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.21.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.0	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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.20.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.20.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.19.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.19.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.18.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.18.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.17.2	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.17.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.17.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.16.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.16.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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-5.15.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.15.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.2	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.14.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.13.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.13.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.12.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.12.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.11.4	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.11.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.11.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-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-ftpfs-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.10.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-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-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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.9.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.3	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.8.2	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.8.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.7.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.7.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.6.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.6.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.4	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.3	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.5.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.4.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.4.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.3.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.3	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.2.2	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.1	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.0	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-5.1.1	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.1.0	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.3	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
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
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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
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
emr-4.9.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.2	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-4.9.1	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-4.8.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.0	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-4.7.4	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.2	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
emr-4.7.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.6.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.5.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
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
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

Etichetta di rilascio di Amazon EMR	Versione di Spark	Componenti installati con Spark
emr-4.1.0	1.5.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.0.0	1.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Apache Sqoop

Apache Sqoop è uno strumento per il trasferimento di dati tra i database Amazon S3, Hadoop, HDFS e RDBMS. Per ulteriori informazioni, consulta il [sito Web di Apache Sqoop](#). Sqoop è incluso in Amazon EMR rilascio 5.0.0 e successivi. I rilasci precedenti includono Sqoop come applicazione sandbox (ambiente di sperimentazione). Per ulteriori informazioni, consulta [Versioni del rilascio di Amazon EMR 4.x](#).

La seguente tabella indica la versione di Sqoop inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Sqoop.

Per la versione dei componenti installati con Sqoop in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Sqoop per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.14.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

La seguente tabella indica la versione di Sqoop inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Sqoop.

Per la versione dei componenti installati con Sqoop in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Sqoop per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.36.1	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

Argomenti

- [Considerazioni riguardanti Sqoop su Amazon EMR](#)
- [Cronologia dei rilasci di Sqoop](#)

Considerazioni riguardanti Sqoop su Amazon EMR

Considera gli elementi seguenti quando si esegue Sqoop su Amazon EMR.

Utilizzo di Sqoop con l'integrazione HCatalog

Sqoop su Amazon EMR supporta l'[integrazione di Sqoop-HCatalog](#). Quando utilizzi Sqoop per scrivere l'output in una tabella HCatalog in Amazon S3, disabilita la scrittura diretta Amazon EMR impostando le proprietà `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` su `false`. Per ulteriori informazioni, consulta [Uso di HCatalog](#). Puoi utilizzare i comandi `-D mapred.output.direct.NativeS3FileSystem=false` e `-D mapred.output.direct.EmrFileSystem=false` di Hadoop. Se non si disabilita la scrittura diretta, non si verifica alcun errore, ma la tabella viene creata in Amazon S3 e non viene scritto alcun dato.

Supporto di Sqoop JDBC e database

Per impostazione predefinita, Sqoop ha installato un driver MariaDB e PostgreSQL. Il driver PostgreSQL installato per Sqoop funziona solo per PostgreSQL 8.4. Per installare un altro set di connettori JDBC per Sqoop, connettersi al nodo master del cluster e installarli in `/usr/lib/sqoop/lib`. Di seguito sono elencati i collegamenti per vari connettori JDBC:

- MariaDB: [Informazioni sul connettore MariaDB/J](#).
- PostgreSQL: [PostgreSQL JDBC Driver](#).
- SQLServer: [Scarica Microsoft JDBC Driver per SQL Server](#).
- MySQL: [Scarica il connettore/J](#)
- Oracle: [Ottieni i driver Oracle JDBC e UCP da Oracle Maven Repository](#)

I database supportati da Sqoop sono elencati nel seguente url, http://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html#_supported_databases, dove *version* è la versione di Sqoop che si sta utilizzando, ad esempio 1.4.6. Se la stringa di connessione JDBC non corrisponde a quelle dell'elenco, è necessario specificare un driver.

Ad esempio, è possibile esportare in una tabella del database Amazon Redshift con il seguente comando (per JDBC 4.1):

```
sqoop export --connect jdbc:redshift://$MYREDSHIFTHOST:5439/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver com.amazon.redshift.jdbc41.Driver --
username master --password Mymasterpass1
```

È possibile utilizzare entrambe le stringhe di connessione MariaDB e MySQL, ma se si specifica la stringa di connessione MariaDB, è necessario specificare il driver:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Se si utilizza la crittografia Secure Socket Layer per accedere al database, è necessario utilizzare un URI JDBC come nell'esempio seguente di esportazione Sqoop:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb?
verifyServerCertificate=false&useSSL=true&requireSSL=true --table mysqoopexport
```

```
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Per ulteriori informazioni sulla crittografia SSL in RDS, consulta [Utilizzo di SSL per crittografare una connessione a un'istanza database](#) nella Guida per l'utente di Amazon RDS.

Per ulteriori informazioni, consulta la documentazione di [Apache Sqoop](#).

Cronologia dei rilasci di Sqoop

La seguente tabella indica la versione di Sqoop inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Sqoop

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.13.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.12.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
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

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.11.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.10.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.10.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.9.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.9.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.8.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.8.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.7.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.36.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.6.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.35.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.5.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.4.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.3.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.3.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.2.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-6.1.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.34.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.33.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.33.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.32.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.32.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.31.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.31.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.30.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.29.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.28.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.28.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.27.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.27.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.26.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.25.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.24.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.24.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.23.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.23.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.22.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.21.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.21.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.20.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.20.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.17.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.16.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.16.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.15.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.14.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.13.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.13.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.12.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.12.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.11.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.11.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.10.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.10.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.9.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.5.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.5.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.3.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.2.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.2.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.1.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.1.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Etichetta di rilascio di Amazon EMR	Versione Sqoop	Componenti installati con Sqoop
emr-5.0.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

TensorFlow

TensorFlow è una libreria matematica simbolica open source per applicazioni di intelligenza artificiale e di deep learning. Per ulteriori informazioni, consulta il [sito Web TensorFlow](#). TensorFlow è disponibile con Amazon EMR versione 5.17.0 e successive.

La seguente tabella indica la versione di TensorFlow inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con TensorFlow.

Per la versione dei componenti installati con TensorFlow in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di TensorFlow per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.14.0	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

La seguente tabella indica la versione di TensorFlow inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con TensorFlow.

Per la versione dei componenti installati con TensorFlow in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di TensorFlow per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.36.1	TensorFlow 2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Build di TensorFlow per tipo di istanza Amazon EC2

Amazon EMR utilizza diverse build della libreria TensorFlow a seconda del tipo di istanza scelto per il cluster. La tabella seguente elenca le build in base al tipo di istanza.

Tipi di istanza EC2	Build di TensorFlow
M5 e C5	Tensorflow 1.9.0 con l'ottimizzazione Intel MKL
P2	Tensorflow 1.9.0 con CUDA 9.2, cuDNN 7.1
P3	Tensorflow 1.9.0 con CUDA 9.2, cuDNN 7.1, NCCL 2.2.13 Nvidia NCCL è disponibile solo per le istanze P3. Contratto di licenza con l'utente finale (EULA): utilizzando i componenti Nvidia su Amazon EMR, accetti i termini e le condizioni illustrate nell' EULA del prodotto .
Tutti gli altri	Tensorflow 1.9.0

Sicurezza

Oltre a seguire le indicazioni in [Utilizzo sicuro di TensorFlow](#), ti consigliamo di avviare il cluster in una sottorete privata per limitare l'accesso a origini affidabili. Per ulteriori informazioni, consulta [Opzioni di Amazon VPC](#) nella Guida alla gestione di Amazon EMR.

Utilizzo di TensorBoard

TensorBoard è una suite di strumenti di visualizzazione per i programmi TensorFlow. Per ulteriori informazioni, consulta [TensorBoard: Visualized learning \(TensorBoard: apprendimento visivo\)](#) sul sito Web di Tensorflow.

Per utilizzare TensorBoard con Amazon EMR, è necessario avviare TensorBoard nel nodo master del cluster.

Utilizzo di Tensorboard con Tensorflow su Amazon EMR

1. Connettersi al nodo master del cluster utilizzando SSH. Per ulteriori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
2. Digitare il seguente comando per avviare Tensorboard sul nodo master. Sostituisci */my/log/directory* con una directory del nodo master in cui sono stati generati e memorizzati i dati di riepilogo utilizzando un programma di scrittura di dati di riepiloghi.

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
```

Per impostazione predefinita, il nodo master ospita TensorBoard utilizzando la porta 6006 e il nome DNS master pubblico. Una volta avviato TensorBoard, l'output della riga di comando mostra l'URL che può essere utilizzato per connettersi a TensorBoard, come nell'esempio seguente:

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. Configurazione dell'accesso alle interfacce Web sul nodo master di client affidabili. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce Web ospitate nei cluster Amazon EMR](#) nella Guida alla gestione di Amazon EMR.
4. Apri TensorBoard a `http://master-public-dns-name:6006`.

Cronologia delle versioni di TensorFlow

La seguente tabella indica la versione di TensorFlow inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di TensorFlow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.14.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.13.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.12.0	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-6.11.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.11.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con 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
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-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.9.0	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.8.1	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.8.0	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.7.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.6.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.35.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.5.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.4.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.3.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.3.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-6.2.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.1.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con 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-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.0.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, 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-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.34.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.33.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.32.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.31.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.31.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.2	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.30.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.29.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.28.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.28.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.27.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.27.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.26.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.25.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.24.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.24.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.23.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.22.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.2	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.21.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.20.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.20.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.19.1	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.0	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.18.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.18.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.2	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Etichetta di rilascio di Amazon EMR	Versione TensorFlow	Componenti installati con TensorFlow
emr-5.17.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Apache Tez

Apache Tez è un framework per la creazione di un grafo aciclico orientato (DAG) complesso delle attività per l'elaborazione di dati. È possibile utilizzarlo come alternativa a Hadoop MapReduce per alcuni casi d'uso. Ad esempio, puoi eseguire i flussi di lavoro Pig e Hive utilizzando Hadoop MapReduce oppure puoi usare Tez come motore di esecuzione. Per ulteriori informazioni, consulta la pagina <https://tez.apache.org/>. Le versioni 4.7.0 e successive di Amazon EMR includono Tez.

La seguente tabella indica la versione di Tez inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Tez.

Per la versione dei componenti installati con Tez in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Tez per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.14.0	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

La seguente tabella indica la versione di Tez inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Tez.

Per la versione dei componenti installati con Tez in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Tez per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.36.1	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

Argomenti

- [Creazione di un cluster con Tez](#)
- [Configurazione di Tez](#)
- [Interfaccia utente Web di Tez](#)
- [Timeline Server](#)
- [Cronologia dei rilasci di Tez](#)

Creazione di un cluster con Tez

Per installare Tez, scegli Apache Tez come applicazione durante la creazione del cluster.

Per creare un cluster con Tez installato usando la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta [Utilizzo della vecchia console](#).
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Configurazione software, seleziona una Versione di emr-4.7.0 o successiva.
4. Seleziona Tez insieme ad altre applicazioni per l'installazione da parte di Amazon EMR.
5. Seleziona le altre opzioni come richiesto, quindi scegli Create cluster (Crea cluster).

Per creare un cluster con Tez usando l'AWS CLI

- Utilizzare il comando `create-cluster` con l'opzione `--applications` per specificare Tez. L'esempio seguente consente di creare un cluster con Tez installato.

Note

I caratteri di continuazione della riga Linux (`\`) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (`^`).

```
aws emr create-cluster --name "Cluster with Tez" --release-label emr-5.36.1 \  
--applications Name=Tez --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configurazione di Tez

Puoi personalizzare Tez impostando i valori tramite la classificazione di configurazione `tez-site`, che consente di configurare le impostazioni nel file di configurazione `tez-site.xml`. Per ulteriori informazioni, consulta [TezConfiguration](#) nella documentazione di Apache Tez. Per modificare Hive o Pig perché impieghino il motore di esecuzione Tez, utilizza le classificazioni di configurazione `hive-site` e `pig-properties` in base alle esigenze. Di seguito sono riportati alcuni esempi.

Example Esempio: Personalizzazione del livello di registrazione root di Tez e impostazione di Tez come motore di esecuzione per Hive e Pig

Il comando `create-cluster` di esempio riportato di seguito crea un cluster con Tez, Hive e Pig installati. Il comando fa riferimento a un file archiviato in Amazon S3, `myConfig.json`, il quale specifica le proprietà per la classificazione `tez-site`, che imposta `tez.am.log.level` su `DEBUG` e definisce Tez come motore di esecuzione per Hive e Pig utilizzando le configurazioni di classificazione `hive-site` e `pig-properties`.

Note

I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
aws emr create-cluster --release-label emr-5.36.1 \  
--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
```

Contenuti di esempio di `myConfig.json` sono mostrati di seguito.

```
[  
  {  
    "Classification": "tez-site",  
    "Properties": {  
      "tez.am.log.level": "DEBUG"  
    }  
  },  
  {  
    "Classification": "hive-site",  
    "Properties": {  
      "hive.execution.engine": "tez"  
    }  
  },  
  {  
    "Classification": "pig-properties",  
    "Properties": {  
      "exectype": "tez"  
    }  
  }  
]
```

Note

Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze

in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta [Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione](#).

Interfaccia utente Web di Tez

Tez dispone di una propria interfaccia utente Web. Per visualizzare l'interfaccia utente Web, consulta il seguente URL.

```
http://masterDNS:8080/tez-ui
```

Per abilitare la scheda Hive Queries nell'interfaccia utente Web di Tez, impostare la seguente configurazione.

```
[
  {
    "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"
    }
  }
]
```

È inoltre possibile visualizzare i dettagli dell'interfaccia utente dell'applicazione Tez, Spark e YARN utilizzando i collegamenti nella scheda Application user interfaces (Interfacce utente delle applicazioni) della pagina dei dettagli di un cluster nella console. Le interfacce utente delle applicazioni Amazon EMR sono ospitate fuori dal cluster e sono disponibili dopo la chiusura del cluster. Non richiedono la configurazione di una connessione SSH o di un proxy Web, semplificando la risoluzione dei problemi e l'analisi dei processi attivi e della cronologia dei processi.

Per ulteriori informazioni, consulta [Visualizzazione della cronologia dell'applicazione](#) nella Guida alla gestione di Amazon EMR.

Timeline Server

Quando Tez è installato, il Timeline Server di YARN è configurato per l'esecuzione. Per visualizzare i processi inviati tramite i motori di esecuzione Tez o MapReduce utilizzando il Timeline Server, accedi all'interfaccia utente Web con l'URL `http://master-public-DNS:8188`. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce Web ospitate nei cluster Amazon EMR](#) nella Guida per la gestione di Amazon EMR.

Cronologia dei rilasci di Tez

La seguente tabella indica la versione di Tez inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio in Versioni del rilascio di Amazon EMR 5.x o Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Tez

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
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
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-resour

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
		cemanager, 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

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.11.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.9.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.9.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.8.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.8.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.7.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.36.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.36.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.6.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.35.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.5.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.4.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.3.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.3.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.1.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.1.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.0.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-6.0.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.34.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.33.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.33.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.31.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.31.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.30.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.29.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.28.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.28.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.27.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.27.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.26.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.25.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.24.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.24.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.23.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.23.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.21.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.20.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.20.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.18.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.18.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.17.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.16.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.14.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.13.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.13.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.12.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.11.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.11.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.10.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.6.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.6.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.5.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.5.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.3.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.2.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.2.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-5.0.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn
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

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-4.9.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
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

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
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
emr-4.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-4.8.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
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

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
emr-4.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-4.7.2	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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn

Etichetta di rilascio di Amazon EMR	Versione di Tez	Componenti installati con Tez
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

Note sul rilascio di Tez per versione

Argomenti

- [Amazon EMR 6.14.0: note di rilascio per Tez](#)
- [Amazon EMR 6.13.0: note di rilascio per Tez](#)
- [Amazon EMR 6.12.0: note di rilascio per Tez](#)
- [Amazon EMR 6.11.0: note di rilascio per Tez](#)
- [Amazon EMR 6.10.0: note di rilascio per Tez](#)
- [Amazon EMR 6.9.0 - Note sul rilascio per Tez](#)
- [Amazon EMR 6.8.0 - Note sul rilascio di Tez](#)
- [Amazon EMR 6.7.0 - Note sul rilascio di Tez](#)

- [Amazon EMR 6.6.0 - Note sul rilascio di Tez](#)

Amazon EMR 6.14.0: note di rilascio per Tez

Amazon EMR 6.14.0: modifiche di Tez

Tipo	Descrizione
Miglioramenti	Aggiorna la versione TLS in Tez alla 1.2

Amazon EMR 6.13.0: note di rilascio per Tez

Amazon EMR 6.13.0: modifiche di Tez

Tipo	Descrizione
Correzioni di bug	REVERT TEZ-4295 : impossibile decomprimere i dati. La durata del buffer è troppo breve.
Correzioni di bug	REVERT TEZ-4302 : <code>NullPointerException</code> in <code>CodecUtils</code> con <code>GzipCodec</code> .
Correzioni di bug	REVERT TEZ-4234 : il compressore può causare <code>IllegalArgumentException</code> in <code>Buffer.limit</code> laddove il limite supera la capacità.
Correzioni di bug	REVERT TEZ-4135 : migliora l'allocazione della memoria durante l'esecuzione di letture in memoria.

Amazon EMR 6.12.0: note di rilascio per Tez

Amazon EMR 6.12.0: modifiche di Tez

Tipo	Descrizione
Miglioramenti	Aggiunto il supporto per JDK 11 e JDK 17 Runtime
Correzioni di bug	TEZ-4492 : aggiornamento di Bowerrc per utilizzare il mirroring bower.herokuapp ed evitare il problema CERT_EXPIRE del registro Bower (BOWER-2608)
Upgrade	Surefire aggiornato a 3.0.0-M7

Amazon EMR 6.11.0: note di rilascio per Tez

Amazon EMR 6.11.0: modifiche di Tez

Tipo	Descrizione
Bug	Corretto l'errore relativo alla transizione dello stato dei vertici non valida durante la pulizia a livello di vertice dei dati shuffle
Bug	Corretto l'errore relativo alla pulizia a livello di DAG o vertice dei dati shuffle non funzionanti
Miglioramenti	Abilitazione di <code>tez.am.dag.cleanup.on.completion</code> per impostazione predefinita per cancellare i dati shuffle dei DAG completati

Amazon EMR 6.10.0: note di rilascio per Tez

Amazon EMR 6.10.0: modifiche di Tez

Tipo	Descrizione
Caratteristica	Abilitazione di <code>tez.runtime.transfer.data-via-events.enabled</code> per impostazione predefinita
Backport	TEZ-4450 : correzione dell'errore di recupero dei dati shuffle quando questi ultimi vengono trasferiti tramite eventi di spostamento dei dati
Backport	TEZ-4460 : correzione dell'errore di timeout di lettura nel recupero dei dati shuffle da Tez Shuffle Handler
Backport	TEZ-4455 : aggiunta di LoggingHandler nella pipeline ShuffleHandler per una possibilità di debug
Bug	Correzione dell'errore relativo all'attività Tez che si blocca a intermittenza quando la priorità dell'attività è abilitata

Amazon EMR 6.9.0 - Note sul rilascio per Tez

Amazon EMR 6.9.0 - Modifiche a Tez

Tipo	Descrizione
Upgrade	Tez è stato aggiornato a 0.10.2. Per ulteriori informazioni, consulta il log delle modifiche di Apache Tez 0.10.2 .
Upgrade	Hadoop è stato aggiornato a 3.3.3.

Tipo	Descrizione
Bug	Disabilita <code>tez.runtime.transfer.data-via-events.enabled</code> per impostazione predefinita a causa di TEZ-4450 .

Amazon EMR 6.8.0 - Note sul rilascio di Tez

Amazon EMR 6.8.0 - Modifiche a Tez

Tipo	Descrizione
Backport	TEZ-3363 : elimina i dati intermedi a livello di vertice per Shuffle Handler
Backport	TEZ-4129 : elimina i dati dei tentativi intermedi per i tentativi falliti per Shuffle Hand
Backport	TEZ-4430 : risolto il problema con la proprietà <code>tez.task.launch.cmd-opts</code> che non funzionava

Amazon EMR 6.7.0 - Note sul rilascio di Tez

Amazon EMR 6.7.0 - Modifiche a Tez

Tipo	Descrizione
Backport	TEZ-4403 : aggiorna la versione SLF4J alla 1.7.36
Backport	TEZ-4405 : sostituisci log4j 1.x con reload4j
Backport	TEZ-4411 : errore della build Tez: FileSaver.js non trovato

Amazon EMR 6.6.0 - Note sul rilascio di Tez

Amazon EMR 6.6.0 - Modifiche a Tez

Tipo	Descrizione
Backport	TEZ-3918 : corretta la proprietà tez.task.log.level non funzionante.
Backport	TEZ-4353 : aggiornato commons-io a 2.8.0.
Backport	TEZ-4114 : rimossa la dipendenza jetty diretta da tez.
Backport	TEZ-4323 : i jar jetty sono stati rimossi dal pacchetto dist con TEZ-4114.

Apache Zeppelin

Utilizzare Apache Zeppelin come notebook per l'esplorazione interattiva dei dati. Per ulteriori informazioni su Zeppelin, visita la pagina Web all'indirizzo <https://zeppelin.apache.org/>. Zeppelin è incluso in Amazon EMR rilascio 5.0.0 e successivi. I rilasci precedenti includono Zeppelin come applicazione sandbox (ambiente di sperimentazione). Per ulteriori informazioni, consulta [Versioni del rilascio di Amazon EMR 4.x](#).

Per accedere all'interfaccia web Zeppelin, impostare un tunnel SSH al nodo master e una connessione proxy. Per ulteriori informazioni, consulta [Visualizzazione delle interfacce Web ospitate su cluster Amazon EMR](#).

La seguente tabella indica la versione di Zeppelin inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con Zeppelin.

Per la versione dei componenti installati con Zeppelin in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Zeppelin per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.14.0	Zeppelin 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

La seguente tabella indica la versione di Zeppelin inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con Zeppelin.

Per la versione dei componenti installati con Zeppelin in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di Zeppelin per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.36.1	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

Argomenti

- [Considerazioni sull'utilizzo di Zeppelin su Amazon EMR](#)
- [Cronologia dei rilasci di Zeppelin](#)

Considerazioni sull'utilizzo di Zeppelin su Amazon EMR

- Connettersi a Zeppelin utilizzando lo stesso [metodo di tunneling SSH](#) per connettersi ad altri server Web sul nodo master. Il server Zeppelin si trova nella porta 8890.
- Zeppelin su Amazon EMR versioni 5.0.0 e superiori supporta l'[autenticazione Shiro](#).

- Zeppelin su Amazon EMR versione 5.8.0 e successive supporta l'utilizzo di AWS Glue Data Catalog come metastore per Spark SQL. Per ulteriori informazioni, consulta la sezione relativa all'[Utilizzo di AWS Glue Data Catalog come metastore per Spark SQL](#).
- Zeppelin non utilizza alcune delle impostazioni definite nel file di configurazione del cluster `spark-defaults.conf`, anche se istruisce YARN di assegnare gli executor in modo dinamico se `spark.dynamicAllocation.enabled` è impostato su `true`. È necessario impostare le impostazioni esecutore, ad esempio memoria e core, utilizzando la scheda Interpreter (Interprete) di Zeppelin, quindi riavviare l'interprete affinché possa essere utilizzato.
- Le versioni 6.10.0 e successive di Amazon EMR supportano l'integrazione di Apache Zeppelin con Apache Flink. Per ulteriori informazioni, consulta [Utilizzo dei processi Flink di Zeppelin in Amazon EMR](#).
- Zeppelin su Amazon EMR non supporta l'interprete SparkR.

Cronologia dei rilasci di Zeppelin

La tabella seguente indica la versione di Zeppelin inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di Zeppelin

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.14.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server,

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
		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-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-6.12.0	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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.7.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.1	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.36.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.6.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.35.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.5.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.4.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.3.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.3.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.2.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.2.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.1.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.1.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.0.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-6.0.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.34.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.33.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.33.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.32.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.32.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.31.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.31.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.30.2	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.30.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.30.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.29.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.28.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.28.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.27.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.27.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.26.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.25.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.24.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.24.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.23.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.23.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.22.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.2	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.21.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.20.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.20.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.19.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.19.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.18.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.18.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.17.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.17.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.17.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.16.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.16.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.15.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.15.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.14.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.13.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.13.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.12.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.12.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.11.4	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.11.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.11.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.10.1	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.10.0	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.9.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.9.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.3	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.8.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.7.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.7.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.6.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.6.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.4	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.5.3	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.2	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.5.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.4.1	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.4.0	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.3.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.3.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.2.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
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
emr-5.1.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.1.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Etichetta di rilascio di Amazon EMR	Versione di Zeppelin	Componenti installati con Zeppelin
emr-5.0.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Apache ZooKeeper

Apache ZooKeeper è un servizio centralizzato concepito per la manutenzione delle informazioni di configurazione, i servizi di denominazione, la sincronizzazione distribuita e l'erogazione di servizi di gruppo. Per ulteriori informazioni su ZooKeeper, visita la pagina Web all'indirizzo <http://zookeeper.apache.org/>.

La seguente tabella indica la versione di ZooKeeper inclusa nell'ultimo rilascio della serie Amazon EMR 6.x insieme ai componenti che Amazon EMR installa con ZooKeeper.

Per la versione dei componenti installati con ZooKeeper in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di ZooKeeper per emr-6.14.0

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-6.14.0	ZooKeeper 3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

La seguente tabella indica la versione di ZooKeeper inclusa nell'ultimo rilascio della serie Amazon EMR 5.x insieme ai componenti che Amazon EMR installa con ZooKeeper.

Per la versione dei componenti installati con ZooKeeper in questo rilascio, consulta la sezione [Versioni dei componenti del rilascio 6.14.0](#).

Informazioni sulla versione di ZooKeeper per emr-5.36.1

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.36.1	ZooKeeper 3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Argomenti

- [Cronologia dei rilasci di ZooKeeper](#)

Cronologia dei rilasci di ZooKeeper

La seguente tabella indica la versione di ZooKeeper inclusa in ogni rilascio di Amazon EMR e i componenti installati con l'applicazione. Per le versioni dei componenti in ogni rilascio, consulta la sezione [Versione componente per il rilascio](#) in [Versioni del rilascio di Amazon EMR 5.x](#) o [Versioni del rilascio di Amazon EMR 4.x](#).

Informazioni sulla versione di ZooKeeper

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.13.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.12.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-6.11.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.11.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.10.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-6.8.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.8.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.7.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.36.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.6.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.35.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.5.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.4.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-6.3.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.3.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.2.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-6.2.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.33.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.33.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.32.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.32.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.30.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.27.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.27.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.26.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.25.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.23.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.23.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.22.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.21.2	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.20.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.20.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.19.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.19.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.17.2	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.16.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.16.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.15.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.15.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.14.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.12.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.12.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.11.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.9.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.8.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.7.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.7.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.5.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con ZooKeeper
emr-5.4.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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

Etichetta di rilascio di Amazon EMR	Versione di ZooKeeper	Componenti installati con 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-5.0.0	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

Connettori e utility

Amazon EMR fornisce diversi connettori e utility per accedere ad altri servizi AWS come origini dati. In genere, puoi accedere ai dati in questi servizi dall'interno di un programma. Ad esempio, puoi specificare un flusso Kinesis in una query Hive, in uno script Pig o in un'applicazione MapReduce e quindi operare su tali dati.

Argomenti

- [Esportazione, importazione, esecuzione di query e unione di tabelle in DynamoDB con Amazon EMR](#)
- [Kinesis](#)
- [S3DistCp \(s3-dist-cp\)](#)
- [Eliminazione dopo attività S3DistCp non riuscite](#)

Esportazione, importazione, esecuzione di query e unione di tabelle in DynamoDB con Amazon EMR

Note

Il connettore Amazon EMR-DynamoDB è open source su GitHub. Per ulteriori informazioni, consulta <https://github.com/awslabs/emr-dynamodb-connector>.

DynamoDB è un servizio di database NoSQL interamente gestito che combina prestazioni elevate e prevedibili con una scalabilità continua. Gli sviluppatori possono creare una tabella di database e accrescerne la richiesta di traffico o lo storage senza limiti. DynamoDB distribuisce automaticamente i dati e il traffico per la tabella su un numero sufficiente di server per gestire la capacità di richieste specificata dal cliente e la quantità di dati archiviati, garantendo al contempo prestazioni rapide e costanti. Utilizzando Amazon EMR e Hive è possibile elaborare in modo rapido ed efficiente grandi quantità di dati, ad esempio i dati archiviati in DynamoDB. Per ulteriori informazioni su DynamoDB, consulta [Guida per gli sviluppatori di Amazon DynamoDB](#).

Apache Hive è un livello software che è possibile utilizzare per eseguire query su cluster MapReduce utilizzando un linguaggio di query semplificato, simile a SQL, denominato HiveQL, e che viene

eseguito su architettura Hadoop. Per ulteriori informazioni su Hive e HiveQL, consulta il [Manuale sul linguaggio HiveQL](#). Per ulteriori informazioni su Hive e Amazon EMR, consulta [Apache Hive](#).

È possibile utilizzare Amazon EMR con una versione personalizzata di Hive, che include la connettività a DynamoDB per eseguire operazioni sui dati archiviati in DynamoDB:

- Caricamento di dati DynamoDB nel file system distribuito Hadoop (HDFS) e loro utilizzo come input in un cluster Amazon EMR.
- Esecuzione di query in diretta sui dati DynamoDB utilizzando istruzioni di tipo SQL (HiveQL).
- Unione di dati archiviati in DynamoDB ed esportazione o esecuzione di query sui dati uniti.
- Esportazione dei dati archiviati in DynamoDB su Amazon S3.
- Importazione dei dati archiviati in Amazon S3 su DynamoDB.

Note

Il connettore Amazon EMR-DynamoDB non supporta i cluster configurati per utilizzare [l'autenticazione Kerberos](#).

Potrai eseguire ognuna delle attività seguenti avviando un cluster Amazon EMR, specificando il percorso dei dati in DynamoDB ed eseguendo comandi Hive per gestire i dati in DynamoDB.

Un cluster Amazon EMR può essere avviato in diversi modi: tramite la console di Amazon EMR, l'interfaccia a riga di comando (CLI) oppure programmando il cluster tramite un SDK AWS o l'API Amazon EMR. È anche possibile scegliere se eseguire un cluster Hive interattivamente o da uno script. In questa sezione illustreremo come avviare un cluster Hive interattivo dalla console di Amazon EMR e dalla CLI.

L'utilizzo interattivo di Hive è perfetto per testare le prestazioni delle query e ottimizzare l'applicazione. Dopo aver creato un set di comandi Hive che verrà eseguito periodicamente, valuta la possibilità di creare uno script Hive che Amazon EMR può eseguire senza interventi da parte tua.

Warning

Le operazioni di lettura o scrittura di Amazon EMR su una tabella DynamoDB vengono conteggiate ai fini della velocità effettiva assegnata stabilita, aumentando potenzialmente la frequenza delle eccezioni alla velocità effettiva assegnata. Per le richieste di grandi dimensioni, Amazon EMR implementa tentativi con backoff esponenziale per gestire il

carico di richieste sulla tabella DynamoDB. L'esecuzione contemporanea di processi Amazon EMR con altro traffico potrebbe causare il superamento del livello di velocità effettiva assegnata allocata. È possibile monitorare questa situazione selezionando il parametro `ThrottleRequests` in Amazon CloudWatch. Se il carico di richieste è troppo elevato, è possibile riavviare il cluster e impostare [Impostazione della percentuale di lettura](#) o [Impostazione della percentuale di scrittura](#) su un valore inferiore per limitare le operazioni Amazon EMR. Per ulteriori informazioni sulle impostazioni di velocità effettiva di DynamoDB, consulta [Velocità effettiva assegnata](#).

Se una tabella è configurata per la [Modalità on demand](#), è necessario ripristinare la tabella in modalità di provisioning prima di eseguire un'operazione di esportazione o importazione. Le pipeline necessitano di un rapporto di velocità effettiva per calcolare le risorse da utilizzare da una tabella DynamoDB. La modalità on demand rimuove la velocità effettiva di provisioning. Per eseguire il provisioning della capacità di velocità effettiva, puoi utilizzare i parametri di Amazon CloudWatch Events per valutare la velocità effettiva aggregata utilizzata da una tabella.

Argomenti

- [Impostazione di una tabella Hive per l'esecuzione dei comandi Hive](#)
- [Esempi di comandi Hive per l'esportazione, l'importazione e l'esecuzione di query sui dati in DynamoDB](#)
- [Ottimizzazione delle prestazioni per le operazioni Amazon EMR in DynamoDB](#)

Impostazione di una tabella Hive per l'esecuzione dei comandi Hive

Apache Hive è un'applicazione di data warehouse che è possibile usare per eseguire query sui dati contenuti nei cluster Amazon EMR utilizzando un linguaggio di tipo SQL. Per ulteriori informazioni su Hive, visita la pagina Web all'indirizzo <http://hive.apache.org/>.

La procedura seguente presuppone che sia già stato creato un cluster e specificata una coppia di chiavi Amazon EC2. Per ulteriori informazioni su come iniziare a creare cluster, consulta [Nozioni di base su Amazon EMR](#) nella Guida alla gestione di Amazon EMR.

Configurazione di Hive per l'uso di MapReduce

Quando usi Hive su Amazon EMR per eseguire query sulle tabelle DynamoDB, possono verificarsi degli errori se Hive utilizza il motore di esecuzione di default, Tez. Per questo motivo, al momento

della creazione di un cluster con Hive che si integra con DynamoDB come descritto in questa sezione, è consigliabile utilizzare una classificazione di configurazione che imposta Hive per l'utilizzo con MapReduce. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

Il seguente snippet mostra la classificazione delle configurazioni e la proprietà da utilizzare per impostare MapReduce come motore di esecuzione per Hive:

```
[
    {
        "Classification": "hive-site",
        "Properties": {
            "hive.execution.engine": "mr"
        }
    }
]
```

Per eseguire i comandi Hive in modo interattivo

1. Connettersi al nodo master. Per maggiori informazioni, consulta [Connessione al nodo master tramite SSH](#) nella Guida alla gestione di Amazon EMR.
2. Nel prompt dei comandi per il nodo master corrente, digitare `hive`.

Viene visualizzato un prompt hive: `hive>`

3. Inserisci un comando Hive che mappi una tabella nell'applicazione Hive ai dati in DynamoDB. Questa tabella agisce come riferimento ai dati archiviati in Amazon DynamoDB: i dati non vengono memorizzati localmente in Hive e le query che utilizzano la tabella sono eseguite sui dati live in DynamoDB. Pertanto, la capacità di lettura o scrittura della tabella viene consumata ogni volta che viene eseguito un comando. Se è prevista l'esecuzione di più comandi Hive sullo stesso set di dati, valutare la possibilità di esportarli.

Di seguito è mostrata la sintassi per la mappatura di una tabella Hive a una tabella DynamoDB.

```
CREATE EXTERNAL TABLE hive_tablename
(hive_column1_name column1_datatype, hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name, hive_column2_name:dynamodb_attribute2_name...")
```

Le tabelle create in Hive da DynamoDB devono essere create come tabelle esterne utilizzando la parola chiave `EXTERNAL`. La differenza tra le tabelle interne ed esterne è che, nel primo caso, i dati vengono eliminati quando viene rimossa la tabella interna. In caso di connessione ad Amazon DynamoDB non si tratta di un comportamento ottimale, quindi sono supportate solo le tabelle esterne.

Ad esempio, il seguente comando Hive crea una tabella denominata `hivetable1` in Hive che fa riferimento alla tabella DynamoDB denominata `dynamoddbtable1`. La tabella DynamoDB `dynamoddbtable1` ha uno schema a chiave primaria hash e di intervallo. L'elemento della chiave hash è `name` (tipo `String`), l'elemento della chiave di intervallo è `year` (tipo `Numeric`) e ciascuna voce ha un valore attributo per `holidays` (tipo `String Set`).

```
CREATE EXTERNAL TABLE hivetable1 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamoddbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

La riga 1 usa l'istruzione HiveQL `CREATE EXTERNAL TABLE`. Per `hivetable1`, è necessario stabilire una colonna per ogni coppia nome-valore dell'attributo nella tabella DynamoDB e fornire il tipo di dati. Questi valori non prevedono distinzione fra maiuscole e minuscole ed è possibile dare alle colonne qualsiasi nome (eccetto le parole riservate).

La riga 2 usa l'istruzione `STORED BY`. Il valore di `STORED BY` è il nome della classe che gestisce la connessione tra Hive e DynamoDB. Deve essere impostato su `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'`.

La riga 3 usa l'istruzione `TBLPROPERTIES` per associare `"hivetable1"` alla tabella e allo schema corretti in DynamoDB. Fornire `TBLPROPERTIES` con i valori per i parametri `dynamodb.table.name` e `dynamodb.column.mapping`. Per questi valori è prevista la distinzione tra maiuscole e minuscole.

Note

Tutti i nomi di attributi DynamoDB per la tabella devono avere colonne corrispondenti nella tabella Hive. Se la mappatura uno-a-uno non esiste, a seconda della versione di Amazon EMR si verificano i seguenti scenari:

- In Amazon EMR 5.27.0 e versioni successive, il connettore dispone di convalide che garantiscono una mappatura uno-a-uno tra i nomi di attributi e le colonne di DynamoDB nella tabella Hive. Se la mappatura uno-a-uno non esiste, si verifica un errore.
- In Amazon EMR 5.26.0 e versioni precedenti, la tabella Hive non contiene la coppia nome-valore proveniente da DynamoDB. Se non vengono mappati gli attributi della chiave primaria DynamoDB, Hive genera un errore. Se non vengono mappati gli attributi della chiave non primaria, non vengono generati errori, ma nella tabella Hive non saranno visualizzati i dati. Se i tipi di dati non corrispondono, il valore è null.

A questo punto puoi iniziare a eseguire operazioni Hive su `hivetable1`. Le query eseguite per `hivetable1` vengono eseguite internamente nella tabella DynamoDB `dynamodbtable1` dell'account DynamoDB, consumando unità di lettura o scrittura a ogni esecuzione.

Quando esegui query Hive su una tabella DynamoDB, devi accertarti di aver assegnato una quantità sufficiente di unità di capacità di lettura.

Ad esempio, supponiamo che tu disponga di 100 unità di capacità di lettura assegnate per la tabella DynamoDB. Questo ti consentirà di eseguire 100 operazioni di lettura, o 409.600 byte, al secondo. Se la tabella contiene 20 GB di dati (21.474.836.480 byte) e la tua query Hive esegue una scansione completa della tabella, puoi stimare la durata dell'esecuzione della query:

$$21.474.836.480 / 409.600 = 52.429 \text{ secondi} = 14,56 \text{ ore}$$

Il solo modo per ridurre il tempo necessario sarebbe modificare le unità di capacità di lettura nella tabella DynamoDB di origine. L'aggiunta di più nodi Amazon EMR non è utile.

Nell'output Hive, la percentuale di completamento viene aggiornata quando terminano uno o più processi del mappatore. Per una tabella DynamoDB di grandi dimensioni con una bassa capacità di lettura assegnata, l'output della percentuale di completamento potrebbe non essere aggiornato per molto tempo; in questo caso, il processo sembrerà essere completo allo 0% per diverse ore. Per uno stato più dettagliato dell'avanzamento del processo, accedi alla console di Amazon EMR, dalla quale potrai visualizzare lo stato delle singole attività del mappatore e statistiche per le letture di dati. Puoi anche accedere all'interfaccia Hadoop sul nodo master e vedere le statistiche di Hadoop. Ti sarà mostrato lo stato della singola attività del mappatore e alcune statistiche di lettura dei dati. Per ulteriori informazioni, consulta i seguenti argomenti:

- [Interfacce Web ospitate sul nodo master](#)
- [Visualizzazione delle interfacce Web di Hadoop](#)

Per ulteriori informazioni sulle istruzioni HiveQL di esempio per eseguire attività quali l'esportazione o l'importazione di dati da DynamoDB e unione di tabelle, consulta [Esempi di comandi Hive per l'esportazione, l'importazione e l'esecuzione di query sui dati in DynamoDB](#).

Per annullare una richiesta Hive

Quando esegui una query Hive, la risposta iniziale del server include il comando per annullare la richiesta. Per annullare la richiesta in qualsiasi momento del processo, utilizza il comando Kill Command (Comando di chiusura) dalla risposta del server.

1. Immettere `Ctrl+C` per chiudere il client a riga di comando.
2. Al prompt della shell, immettere il comando Kill Command (Comando di chiusura) tratto dalla risposta iniziale del server alla richiesta.

In alternativa, è possibile eseguire il comando seguente dalla riga di comando del nodo master per chiudere il processo Hadoop, dove *job-id* è l'identificatore di tale processo e può essere recuperato dall'interfaccia utente Hadoop.

```
hadoop job -kill job-id
```

Tipi di dati per Hive e DynamoDB

La seguente tabella mostra i tipi di dati Hive disponibili, il tipo di DynamoDB predefinito a cui fanno riferimento e i tipi di DynamoDB alternativi a cui possono essere associati.

Tipo Hive	Tipo di DynamoDB predefinito	Tipo/i di DynamoDB alternativo/i
stringa	string (S)	
bigint o double	number (N)	
binary	binary (B)	

Tipo Hive	Tipo di DynamoDB predefinito	Tipo/i di DynamoDB alternativo/i
booleano	boolean (BOOL)	
array	list (L)	number set (NS), string set (SS) o binary set (BS)
map<string,string>	elemento	map (M)
map<string,?>	map (M)	
	null (NULL)	

Se desideri scrivere i dati Hive come un tipo corrispondente di DynamoDB alternativo o se i dati di DynamoDB contengono valori di attributi di un tipo di DynamoDB alternativo, puoi specificare la colonna e il tipo di DynamoDB con il parametro `dynamodb.type.mapping`. L'esempio seguente mostra la sintassi per specificare una mappatura di tipi alternativi.

```
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");
```

Il parametro della mappatura dei tipi è facoltativo e deve essere specificato solo per le colonne che utilizzano tipi alternativi.

Ad esempio, il seguente comando Hive crea una tabella denominata `hivetable2` che fa riferimento alla tabella `dynamodhtable2` di DynamoDB. È simile a `hivetable1`, tranne per il fatto che associa la colonna `col3` al tipo `string set (SS)`.

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
```

```
"dynamodb.type.mapping" = "col3:SS");
```

In Hive, `hivetable1` e `hivetable2` sono identici. Tuttavia, quando i dati di tali tabelle vengono scritti nelle tabelle DynamoDB corrispondenti, `dynamodbtable1` contiene elenchi, mentre `dynamodbtable2` contiene set di stringhe.

Se desideri scrivere valori `null` di Hive come attributi del tipo `null` di DynamoDB, puoi farlo con il parametro `dynamodb.null.serialization`. L'esempio seguente mostra la sintassi per specificare la serializzazione `null`.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,  
hive_column2_name column2_datatype...)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",  
"dynamodb.column.mapping" =  
"hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",  
"dynamodb.null.serialization" = "true");
```

Il parametro di serializzazione `null` è facoltativo e, se non specificato, è impostato su `false`. Nota che gli attributi `null` di DynamoDB sono letti come valori `null` in Hive indipendentemente dalle impostazioni del parametro. Le raccolte Hive con valori `null` possono essere scritte in DynamoDB solo se il parametro di serializzazione `null` è specificato come `true`. In caso contrario, si verifica un errore Hive.

Il tipo `bigint` in Hive è uguale al tipo `long` in Java, mentre il tipo `double` di Hive è uguale al tipo `double` in Java, in termini di precisione. Questo significa che, se disponi di dati numerici memorizzati in DynamoDB con una precisione superiore a quella disponibile nei tipi di dati Hive, l'utilizzo di Hive per esportare, importare o fare riferimento ai dati DynamoDB può portare a una perdita di precisione o a un errore nella query Hive.

Le esportazioni del tipo `binary` da DynamoDB ad Amazon Simple Storage Service (Amazon S3) o HDFS vengono memorizzate come una stringa con codifica Base64. Se importi dati da Amazon S3 o HDFS nel tipo `binary` di DynamoDB, assicurati che siano codificati come stringa Base64.

Opzioni Hive

Puoi impostare le seguenti opzioni di Hive per gestire il trasferimento di dati da Amazon DynamoDB. Queste opzioni persistono nella sessione di Hive corrente. Se chiudi il prompt dei comandi Hive e lo riapri in un secondo momento nel cluster, le impostazioni ritorneranno ai valori predefiniti.

Opzioni Hive	Descrizione
<code>dynamodb.throughput.read.percent</code>	<p>Imposta la percentuale di operazioni di lettura per mantenere il tasso di velocità effettiva assegnato di DynamoDB nell'intervallo allocato per la tabella. Il valore è compreso tra 0.1 e 1.5 inclusi.</p> <p>Il valore di 0,5 è la velocità di lettura di default, il che significa che Hive cercherà di consumare metà delle risorse di throughput assegnate alla lettura nella tabella. L'aumento del valore sopra a 0,5 aumenta la velocità della richiesta di lettura, che invece scende riducendo il valore sotto a 0,5. La velocità di lettura è approssimativa. La velocità di lettura reale varia in base a fattori come la presenza di una distribuzione uniforme delle chiavi in DynamoDB.</p> <p>Se vedi che il throughput assegnato viene frequentemente superato dalle operazioni Hive, oppure se è il traffico live in lettura viene limitato eccessivamente, porta il valore sotto a 0.5. Se disponi di capacità sufficiente e desideri operazioni Hive più veloci, imposta questo valore al di sopra di 0.5. Se ritieni che non ci siano operazioni di input/output inutilizzate disponibili, puoi anche andare oltre impostando il valore fino a 1,5.</p>
<code>dynamodb.throughput.write.percent</code>	<p>Imposta la percentuale di operazioni di scrittura per mantenere il tasso di velocità effettiva assegnato di DynamoDB nell'intervallo allocato per la tabella. Il valore è compreso tra 0.1 e 1.5 inclusi.</p> <p>Il valore di 0,5 è la velocità di scrittura di default, il che significa che Hive cercherà di consumare metà delle risorse di throughput assegnate alla scrittura nella tabella. L'aumento del valore sopra a 0,5 aumenta la velocità delle richieste di scrittura, che invece scende</p>

Opzioni Hive	Descrizione
	<p>riducendo il valore sotto a 0,5. La velocità di scrittura è approssimativa. La velocità di scrittura reale varia in base a fattori come la presenza di una distribuzione uniforme delle chiavi in DynamoDB.</p> <p>Se vedi che il throughput assegnato viene frequentemente superato dalle operazioni Hive, oppure se è il traffico live in scrittura viene limitato eccessivamente, porta il valore sotto a 0.5. Se disponi di capacità sufficiente e desideri operazioni Hive più veloci, imposta questo valore al di sopra di 0.5. Puoi anche andare oltre impostando il valore fino a 1,5, se ritieni che non ci siano operazioni di input/output inutilizzate disponibili o se stai eseguendo il caricamento iniziale dei dati sulla tabella e non c'è ancora traffico live.</p>
<code>dynamodb.endpoint</code>	<p>Specifica l'endpoint per il servizio DynamoDB. Per ulteriori informazioni sugli endpoint DynamoDB disponibili, consulta la pagina relativa a Regioni ed endpoint.</p>
<code>dynamodb.max.map.tasks</code>	<p>Specifica il numero massimo di attività di mappatura durante la lettura di dati da DynamoDB. Questo valore deve essere maggiore o uguale a 1.</p>
<code>dynamodb.retry.duration</code>	<p>Specifica il numero di minuti da utilizzare come durata di timeout per rieseguire i comandi Hive. Questo valore deve essere un numero intero uguale o maggiore di 0. La durata di timeout predefinita è di due minuti.</p>

Queste opzioni vengono impostate utilizzando il comando SET come mostrato nel seguente esempio.

```
SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE TABLE s3_export SELECT *
```

```
FROM hiveTableName;
```

Esempi di comandi Hive per l'esportazione, l'importazione e l'esecuzione di query sui dati in DynamoDB

Nei seguenti esempi vengono utilizzati comandi Hive per eseguire operazioni quali l'esportazione di dati su Amazon S3 o HDFS, l'importazione di dati in DynamoDB, l'unione di tabelle, l'esecuzione di query su tabelle e molto altro.

Le operazioni su una tabella Hive fanno riferimento ai dati archiviati in DynamoDB. I comandi di Hive sono soggetti alle impostazioni di velocità effettiva assegnate alla tabella DynamoDB e i dati recuperati includono i dati scritti nella tabella DynamoDB al momento dell'elaborazione della richiesta di operazione Hive da parte di DynamoDB. Se il processo di recupero dei dati richiede molto tempo, rispetto a quando è iniziato il comando Hive alcuni dei dati restituiti potrebbero essere stati aggiornati in DynamoDB.

I comandi Hive `DROP TABLE` e `CREATE TABLE` agiscono solo sulle tabelle locali in Hive e non creano né eliminano tabelle in DynamoDB. Se la query Hive fa riferimento a una tabella in DynamoDB, quest'ultima deve già esistere prima di eseguire la query. Per ulteriori informazioni sulla creazione e l'eliminazione di tabelle in DynamoDB, consulta [Lavorare con le tabelle in DynamoDB](#) nella Guida per gli sviluppatori di Amazon DynamoDB.

Note

Quando esegui la mappatura di una tabella Hive a una posizione in Amazon S3, evita di mapparla al percorso principale del bucket, `s3://mybucket`, in quanto questo potrebbe causare errori durante le operazioni di scrittura dei dati su Amazon S3 da parte di Hive. Esegui invece la mappatura della tabella a un percorso secondario del bucket, `s3://mybucket/mypath`.

Esportazione di dati da DynamoDB

È possibile utilizzare Hive per esportare i dati da DynamoDB.

Esportazione di una tabella DynamoDB in un bucket Amazon S3

- Crea una tabella Hive che faccia riferimento ai dati archiviati in DynamoDB. A questo punto è possibile chiamare il comando INSERT OVERWRITE per scrivere i dati su una directory esterna. In questo esempio, *s3://bucketname/path/subpath/* è un percorso valido in Amazon S3. Modifica le colonne e i tipi di dati nel comando CREATE (CREA) perché corrispondano ai valori in DynamoDB. È possibile utilizzare questo metodo per creare un archivio dei dati DynamoDB in Amazon S3.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

Esportazione di una tabella DynamoDB in un bucket Amazon S3 utilizzando la formattazione

- Crea una tabella esterna che faccia riferimento a un percorso in Amazon S3. Questo passaggio è mostrato di seguito come *s3_export*. Durante la chiamata CREATE, specificare la formattazione della riga per la tabella. Quindi, quando si utilizza INSERT OVERWRITE per esportare i dati da DynamoDB in *s3_export*, i dati vengono scritti nel formato specificato. In questo esempio, i dati vengono scritti come valori separati da virgola (CSV).

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
```

```
FROM hiveTableName;
```

Esportazione di una tabella DynamoDB in un bucket Amazon S3 senza specificare una mappatura di colonna

- Crea una tabella Hive che faccia riferimento ai dati archiviati in DynamoDB. L'operazione è analoga a quella dell'esempio precedente, salvo che non si specifica una mappatura di colonna. La tabella deve avere esattamente una colonna di tipo `map<string, string>`. Se si crea una tabella EXTERNAL in Amazon S3 è possibile chiamare il comando `INSERT OVERWRITE` per scrivere i dati provenienti da DynamoDB su Amazon S3. È possibile utilizzare questo metodo per creare un archivio dei dati DynamoDB in Amazon S3. Poiché non è presente alcuna mappatura di colonna, non è possibile eseguire query in tabelle esportate in questo modo. L'esportazione dei dati senza specificare una mappatura di colonna è disponibile in Hive versione 0.8.1.5 o successive, supportate sull'AMI Amazon EMR versione 2.2.x e successive.

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "dynamodatabase1");  
  
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;
```

Esportazione di una tabella DynamoDB in un bucket Amazon S3 utilizzando la compressione dei dati

- Hive fornisce diversi codec di compressione che è possibile impostare durante la sessione Hive. In questo modo, i dati esportati vengono compressi nel formato specificato. L'esempio seguente comprime i file esportati utilizzando l'algoritmo Lempel-Ziv-Oberhumer (LZO).

```
SET hive.exec.compress.output=true;  
SET io.seqfile.compression.type=BLOCK;
```



```
SET mapred.output.compression.codec = com.hadoop.compression.lzo.LzopCodec;

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"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;
```

I codec di compressione disponibili sono:

- 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

Esportazione di una tabella DynamoDB in HDFS

- Utilizza il seguente comando Hive, dove *hdfs:///directoryName* è un percorso HDFS valido e *hiveTableName* è una tabella in Hive che fa riferimento a DynamoDB. Questa operazione di esportazione è più veloce dell'esportazione di una tabella DynamoDB su Amazon S3, perché Hive 0.7.1.1 usa HDFS come fase intermedia per l'esportazione di dati su Amazon S3. L'esempio seguente mostra anche come impostare il valore `dynamodb.throughput.read.percent` su 1,0 per aumentare la richiesta di lettura.

```
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;
```

È anche possibile esportare i dati su HDFS utilizzando la formattazione e la compressione come illustrato sopra per l'esportazione ad Amazon S3. Per farlo, è sufficiente sostituire la directory Amazon S3 negli esempi sopra con una directory HDFS.

Per leggere dati di caratteri UTF-8 non stampabili in Hive

- È possibile leggere e scrivere dati di caratteri UTF-8 non stampabili con Hive utilizzando la clausola `STORED AS SEQUENCEFILE` nella creazione di una tabella. SequenceFile è il formato di file binario di Hadoop e necessita appunto di Hadoop per essere letto. Nell'esempio seguente viene illustrato come esportare dati da DynamoDB verso Amazon S3. Puoi utilizzare questa funzionalità per gestire i caratteri in codifica UTF-8 non stampabili.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",  
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");  
  
CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)  
STORED AS SEQUENCEFILE  
LOCATION 's3://bucketname/path/subpath/';  
  
INSERT OVERWRITE TABLE s3_export SELECT *  
FROM hiveTableName;
```

Importazione di dati in DynamoDB

Quando scrivi dati in DynamoDB utilizzando Hive, devi assicurarti che il numero delle unità di capacità di scrittura sia maggiore del numero di mappatori del cluster. Ad esempio, i cluster in esecuzione su istanze EC2 m1.xlarge producono 8 mappatori per istanza. Nel caso di un cluster

che dispone di 10 istanze, ciò significherebbe un totale di 80 mappatori. Se il numero di unità di capacità di scrittura non è superiore al numero di mappatori nel cluster, l'operazione di scrittura Hive potrebbe consumare tutta la velocità effettiva di scrittura o tentare di consumare più velocità effettiva di quanto ne sia stato assegnato. Per ulteriori informazioni sul numero di mappatori creato da ogni tipo di istanza EC2, consulta [Configura Hadoop](#).

Il numero di mappatori in Hadoop è controllato dalle divisioni in entrata. Se le divisioni sono troppo poche, il comando di scrittura potrebbe non essere in grado di consumare tutto il throughput di scrittura disponibile.

Se un elemento con la stessa chiave è presente nella tabella DynamoDB di destinazione, viene sovrascritto. Se un elemento con la stessa chiave non è presente nella tabella DynamoDB di destinazione, viene inserito.

Importazione di una tabella da Amazon S3 a DynamoDB

- È possibile utilizzare Amazon EMR e Hive per scrivere i dati da Amazon S3 a DynamoDB.

```
CREATE EXTERNAL TABLE s3_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

Importazione una tabella da un bucket Amazon S3 a DynamoDB senza specificare una mappatura di colonna

- Crea una tabella EXTERNAL che faccia riferimento a dati archiviati in Amazon S3 ed esportati in precedenza da DynamoDB. Prima dell'importazione, assicurati che la tabella sia presente in DynamoDB e che abbia lo stesso schema di chiave della tabella DynamoDB esportata in precedenza. La tabella deve inoltre avere esattamente una colonna di tipo `map<string, string>`. Se si crea una tabella Hive collegata a DynamoDB è possibile chiamare il comando `INSERT OVERWRITE` per scrivere i dati provenienti da Amazon S3 su DynamoDB. Poiché non

È presente alcuna mappatura di colonna, non è possibile eseguire query in tabelle importate in questo modo. L'importazione dei dati senza specificare una mappatura di colonna è disponibile in Hive versione 0.8.1.5 o successive, supportate sull'AMI Amazon EMR 2.2.3 e successive.

```
CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (item map<string, string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

Importazione di una tabella da HDFS a DynamoDB

- È possibile utilizzare Amazon EMR e Hive per scrivere i dati da HDFS a DynamoDB.

```
CREATE EXTERNAL TABLE hdfs_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'hdfs:///directoryName';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM hdfs_import;
```

Esecuzioni di query sui dati in DynamoDB

Negli esempi seguenti vengono illustrati i diversi modi in cui puoi utilizzare Amazon EMR per eseguire query sui dati archiviati in DynamoDB.

Per trovare il valore maggiore per una colonna mappata (**max**)

- Utilizzare comandi Hive come quelli indicati di seguito. Nel primo comando l'istruzione CREATE (CREA) crea una tabella Hive che faccia riferimento ai dati archiviati in DynamoDB. L'istruzione SELECT usa quindi la tabella per eseguire una query sui dati archiviati in DynamoDB. L'esempio seguente trova l'ordine più grande effettuato da un determinato cliente.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT max(total_cost) from hive_purchases where customerId = 717;
```

Per aggregare i dati utilizzando la clausola **GROUP BY**

- Puoi utilizzare la clausola GROUP BY per raccogliere dati all'interno di più record. Questa clausola viene spesso utilizzata con una funzione di aggregazione come sum, count, min o max. L'esempio seguente restituisce un elenco degli ordini più grandi di clienti che hanno effettuato più di tre ordini.

```
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;
```

Unione di due tabelle DynamoDB

- L'esempio seguente mappa due tabelle Hive ai dati archiviati in DynamoDB. quindi chiama un'operazione di unione tra le due tabelle. L'unione viene calcolata nel cluster e in seguito viene restituita. L'unione non si svolge in DynamoDB. Questo esempio restituisce un elenco di clienti e dei loro acquisti per i clienti che hanno effettuato più di due ordini.

```
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;
```

Per unire due tabelle da origini diverse

- In questo esempio, *Customer_S3* è una tabella Hive che carica un file CSV archiviato in Amazon S3, mentre *hive_purchases* è una tabella che fa riferimento ai dati in DynamoDB. L'esempio seguente unisce i dati dei clienti memorizzati come file CSV in Amazon S3 con i dati degli ordini archiviati in DynamoDB per restituire un set di dati che rappresenta gli ordini effettuati dai clienti il cui nome contiene la parola "Miller".

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
```

```
"dynamodb.column.mapping" =
  "customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE Customer_S3(customerId bigint, customerName string,
  customerAddress array<String>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

Select c.customerId, c.customerName, c.customerAddress from
Customer_S3 c
JOIN hive_purchases p
ON c.customerid=p.customerid
where c.customerName like '%Miller%';
```

Note

In esempi precedenti, le istruzioni CREATE TABLE erano incluse in ciascun esempio per chiarezza e completezza. Se vengono eseguite più operazioni di query o di esportazione su una determinata tabella Hive, è necessario creare la tabella solo una sola volta, all'inizio della sessione Hive.

Ottimizzazione delle prestazioni per le operazioni Amazon EMR in DynamoDB

Le operazioni Amazon EMR su una tabella DynamoDB vengono conteggiate come operazioni di lettura e sono soggette alle impostazioni di velocità effettiva assegnata della tabella. Amazon EMR implementa una propria logica per provare a bilanciare il carico sulla tabella DynamoDB e ridurre al minimo la possibilità di superare la velocità effettiva assegnata. Al termine di ogni query Hive, Amazon EMR restituisce informazioni sul cluster utilizzato per elaborare la query, incluso il numero di volte in cui la velocità effettiva assegnata è stata superata. Queste informazioni, insieme ai parametri di CloudWatch sulla velocità effettiva DynamoDB, sono utili per gestire in modo più efficiente il carico sulla tabella DynamoDB nelle richieste successive.

Nelle operazioni con le tabelle DynamoDB, le prestazioni delle query di Hive sono influenzate dai fattori indicati di seguito.

Unità di capacità di lettura assegnate

Quando esegui query Hive su una tabella DynamoDB, devi accertarti di aver assegnato una quantità sufficiente di unità di capacità di lettura.

Ad esempio, supponiamo che tu disponga di 100 unità di capacità di lettura assegnate per la tabella DynamoDB. Questo ti consentirà di eseguire 100 operazioni di lettura, o 409.600 byte, al secondo. Se la tabella contiene 20 GB di dati (21.474.836.480 byte) e la tua query Hive esegue una scansione completa della tabella, puoi stimare la durata dell'esecuzione della query:

$$21.474.836.480 / 409.600 = 52.429 \text{ secondi} = 14,56 \text{ ore}$$

Il solo modo per ridurre il tempo necessario sarebbe modificare le unità di capacità di lettura nella tabella DynamoDB di origine. L'aggiunta di più nodi al cluster Amazon EMR non è utile.

Nell'output Hive, la percentuale di completamento viene aggiornata quando terminano uno o più processi del mappatore. Per una tabella DynamoDB di grandi dimensioni con una bassa capacità di lettura assegnata, l'output della percentuale di completamento potrebbe non essere aggiornato per molto tempo; in questo caso, il processo sembrerà essere completo allo 0% per diverse ore. Per uno stato più dettagliato dell'avanzamento del processo, accedi alla console di Amazon EMR, dalla quale potrai visualizzare lo stato delle singole attività del mappatore e statistiche per le letture di dati.

Puoi anche accedere all'interfaccia Hadoop sul nodo master e vedere le statistiche di Hadoop. Ti sarà mostrato lo stato della singola attività del mappatore e alcune statistiche di lettura dei dati. Per ulteriori informazioni, consulta la sezione relativa alle [Interfacce Web ospitate sul nodo master](#) nella Guida alla gestione di Amazon EMR.

Impostazione della percentuale di lettura

Per impostazione predefinita, Amazon EMR gestisce il carico di richieste sulla tabella DynamoDB, in base alla velocità effettiva corrente assegnata. Tuttavia, se Amazon EMR restituisce informazioni sul processo in cui è incluso un elevato numero di risposte di superamento della velocità effettiva assegnata, puoi modificare la velocità di lettura di default utilizzando il parametro `dynamodb.throughput.read.percent` durante la configurazione della tabella Hive. Per ulteriori informazioni su come impostare il parametro della percentuale di lettura, consulta [Opzioni Hive](#).

Impostazione della percentuale di scrittura

Per impostazione predefinita, Amazon EMR gestisce il carico di richieste sulla tabella DynamoDB, in base alla velocità effettiva corrente assegnata. Tuttavia, se Amazon EMR restituisce informazioni

sul processo in cui è incluso un elevato numero di risposte di superamento della velocità effettiva assegnata, puoi modificare la velocità di scrittura di default utilizzando il parametro `dynamodb.throughput.write.percent` durante la configurazione della tabella Hive. Per ulteriori informazioni su come impostare il parametro della percentuale di scrittura, consulta [Opzioni Hive](#).

Impostazione della durata dei nuovi tentativi

Per impostazione predefinita, Amazon EMR esegue nuovamente una query Hive se questa non ha restituito un risultato entro due minuti, l'intervallo predefinito per i nuovi tentativi. Puoi modificare questo intervallo impostando il parametro `dynamodb.retry.duration` quando esegui una query Hive. Per ulteriori informazioni su come impostare il parametro della percentuale di scrittura, consulta [Opzioni Hive](#).

Numero di attività di mappatura

I daemon di mappatura che Hadoop avvia per elaborare le richieste di esportazione ed esecuzione di query sui dati archiviati in DynamoDB hanno una velocità di lettura massima di 1 MiB al secondo per limitare la capacità di lettura utilizzata. Se disponi di ulteriore velocità effettiva assegnata disponibile su DynamoDB, puoi migliorare le prestazioni delle operazioni di query ed esportazione di Hive aumentando il numero di daemon del mappatore. A questo scopo, puoi aumentare il numero di istanze EC2 nel cluster o il numero di daemon del mappatore in esecuzione su ciascuna istanza EC2.

Puoi aumentare il numero di istanze EC2 in un cluster arrestando il cluster corrente e riavviandolo con un numero maggiore di istanze EC2. Puoi specificare il numero di istanze EC2 nella finestra di dialogo Configure EC2 Instances (Configura istanze EC2) se avvii il cluster dalla console di Amazon EMR o con l'opzione `--num-instances` se lo avvii dalla CLI.

Il numero di attività di mappatura eseguite su un'istanza dipende dal tipo di istanza EC2. Per ulteriori informazioni sui tipi di istanze EC2 supportate e sul numero di mappatori fornito da ognuno, consulta [Configurazione attività](#), dove troverai una sezione relativa alla configurazione delle attività per ognuna delle configurazioni supportate.

In alternativa, puoi aumentare il numero dei daemon di mappatura modificando il parametro di configurazione `mapreduce.tasktracker.map.tasks.maximum` di Hadoop con un valore superiore. Questo metodo presenta il vantaggio di offrirti un numero maggiore di mappatori senza aumentare né il numero, né le dimensioni delle istanze EC2, con conseguente risparmio di denaro. Lo svantaggio è che un'impostazione troppo elevata in questo valore può provocare l'esaurimento della memoria nelle istanze EC2 del tuo cluster. Per impostare `mapreduce.tasktracker.map.tasks.maximum`, avvia il cluster e specifica un valore per

`mapreduce.tasktracker.map.tasks.maximum` come proprietà della classificazione di configurazione `mapred-site`. Questo viene mostrato nell'esempio seguente. Per ulteriori informazioni, consulta [Configurazione delle applicazioni](#).

```
{
  "configurations": [
    {
      "classification": "mapred-site",
      "properties": {
        "mapreduce.tasktracker.map.tasks.maximum": "10"
      }
    }
  ]
}
```

Richieste di dati in parallelo

Molteplici richieste di dati, sia da parte di più utenti sia da più applicazioni verso un'unica tabella, possono far esaurire il throughput di lettura assegnato e rallentare le prestazioni.

Durata dei processi

La consistenza dei dati in DynamoDB dipende dall'ordine delle operazioni di lettura e scrittura di ciascun nodo. Quando una query Hive è in avanzamento, un'altra applicazione potrebbe caricare nuovi dati nella tabella DynamoDB oppure modificare o eliminare dati esistenti. In questo caso, i risultati della query Hive potrebbe non riflettere le modifiche effettuate ai dati durante l'esecuzione della query.

Evitare di superare la velocità effettiva

Quando esegui query Hive su DynamoDB, fai attenzione a non superare la velocità effettiva assegnata, perché in questo modo si esaurisce la capacità necessaria per le chiamate dell'applicazione a DynamoDB : :Get. Per verificare che non sia questo il caso, è necessario monitorare regolarmente il volume di lettura e il throttling sulle chiamate delle applicazioni a DynamoDB : :Get controllando i log e monitorando i parametri in Amazon CloudWatch.

Ora delle richieste

Le prestazioni possono essere migliorate pianificando query Hive che accedono a una tabella DynamoDB quando la richiesta nella tabella DynamoDB è minore. Ad esempio, se la maggior parte

degli utenti dell'applicazione vive a San Francisco, è possibile scegliere di esportare i dati ogni giorno alle ore 4:00 (PST), quando la maggior parte degli utenti è inattiva e non aggiorna i registri del database DynamoDB.

Tablelle basate sul tempo

Se i dati vengono organizzati in una serie di tabelle DynamoDB basate sul tempo, ad esempio una tabella al giorno, puoi esportare i dati quando la tabella non è più attiva. Puoi utilizzare questa tecnica per eseguire il backup dei dati su Amazon S3 in modo regolare.

Dati archiviati

Se prevedi di eseguire molte query Hive sui dati archiviati in DynamoDB e la tua applicazione è in grado di tollerare i dati archiviati, potrebbe essere una buona idea esportare i dati su HDFS o Amazon S3 ed eseguire le query Hive su una copia dei dati, anziché su DynamoDB. In questo modo, le operazioni di lettura e il throughput assegnato vengono conservati.

Kinesis

I cluster Amazon EMR sono in grado di leggere ed elaborare direttamente i flussi Amazon Kinesis con strumenti familiari nell'ecosistema Hadoop, come Hive, Pig, MapReduce, l'API Hadoop Streaming e Cascading. È anche possibile unire i dati in tempo reale provenienti da Amazon Kinesis con dati esistenti su Amazon S3, Amazon DynamoDB e HDFS in un cluster in esecuzione. I dati possono essere caricati direttamente da Amazon EMR su Amazon S3 o DynamoDB per le attività di post-elaborazione. Per ulteriori informazioni sulle caratteristiche salienti del servizio Amazon Kinesis e sui prezzi, consulta [Amazon Kinesis](#).

Cosa posso fare con l'integrazione di Amazon EMR e Amazon Kinesis?

L'integrazione tra Amazon EMR e Amazon Kinesis semplifica notevolmente alcuni scenari, quali:

- **Analisi dei log di streaming:** puoi analizzare i log Web di streaming per generare ogni pochi minuti un elenco dei 10 errori più diffusi in base alla regione, al browser e al dominio di accesso.
- **Coinvolgimento dei clienti:** puoi scrivere query che uniscano i dati clickstream da Amazon Kinesis con le informazioni sulle campagne pubblicitarie archiviate in una tabella DynamoDB al fine di identificare le categorie di annunci più efficaci tra quelli visualizzati su specifici siti Web.
- **Query interattive ad-hoc:** puoi caricare periodicamente dati dai flussi Amazon Kinesis in HDFS e renderli disponibili come tabella locale Impala per query rapide, interattive e analitiche.

Analisi con checkpoint dei flussi Amazon Kinesis

Gli utenti possono eseguire analisi periodiche in batch dei flussi Amazon Kinesis nelle cosiddette iterazioni. Poiché i registri dei dati di flusso Amazon Kinesis vengono recuperati utilizzando un numero di sequenza, i limiti dell'iterazione sono definiti da numeri di sequenza iniziali e finali che Amazon EMR archivia in una tabella DynamoDB. Ad esempio, al termine di `iteration0`, il numero in sequenza finale viene archiviato in DynamoDB in modo che, all'inizio del processo `iteration1`, sia possibile recuperare i dati successivi dal flusso. Questa mappatura delle iterazioni nei dati di flusso si definisce checkpoint. Per ulteriori informazioni, consulta [Kinesis Connector](#).

Se per un'iterazione sono stati definiti checkpoint e il processo non ha potuto elaborare un'iterazione, Amazon EMR tenta di rielaborare il record in quella iterazione.

I checkpoint consentono di:

- Avviare l'elaborazione dei dati dopo un numero di sequenza elaborato da una precedente query eseguita sullo stesso flusso e nome logico
- Rielaborazione dello stesso batch di dati Kinesis elaborato da una precedente query

Per abilitare i checkpoint, imposta il parametro `kinesis.checkpoint.enabled` su `true` e configura i seguenti parametri:

Impostazione di configurazione	Descrizione
<code>kinesis.checkpoint.metastore.table.name</code>	Nome della tabella DynamoDB in cui saranno archiviate le informazioni dei checkpoint
<code>kinesis.checkpoint.metastore.hash.key.name</code>	Nome chiave hash per la tabella DynamoDB
<code>kinesis.checkpoint.metastore.hash.range.name</code>	Nome chiave range per la tabella DynamoDB
<code>kinesis.checkpoint.logical.name</code>	Un nome logico per l'elaborazione corrente
<code>kinesis.checkpoint.iteration.no</code>	Numero di iterazioni per elaborazione associata al nome logico

Impostazione di configurazione	Descrizione
<code>kinesis.rerun.iteration.without.wait</code>	Valore booleano che indica se un'iterazione non riuscita può essere eseguita nuovamente senza timeout; l'impostazione predefinita è <code>false</code>

Suggerimenti relativi a Provisioned IOPS per le tabelle Amazon DynamoDB

Il connettore Amazon EMR per Amazon Kinesis utilizza il database DynamoDB come database di supporto per il checkpoint dei metadati. Devi creare una tabella in DynamoDB prima di consumare i dati in un flusso Amazon Kinesis con un cluster Amazon EMR in intervalli sui cui è stato eseguito il checkpoint. La tabella deve trovarsi nella stessa regione del cluster Amazon EMR. Di seguito sono riportate le raccomandazioni generali per il numero di IOPS da assegnare alle tabelle DynamoDB, con j che sarà il numero massimo di processi Hadoop (con diverse combinazioni nome logico + numero di iterazione) che possono essere eseguiti simultaneamente e s il numero massimo di shard elaborati da qualsiasi processo:

Per Read Capacity Units (Unità di capacità di lettura): $j*s/5$

Per Write Capacity Units (Unità di capacità di scrittura): $j*s$

Considerazioni sulle prestazioni

La velocità effettiva degli shard Amazon Kinesis è direttamente proporzionale alle dimensioni dell'istanza dei nodi nei cluster Amazon EMR e alle dimensioni dei record nel flusso. Consigliamo di utilizzare l'istanza `m5.xlarge` o istanze più grandi sui nodi master e principali.

Pianificazione dell'analisi Amazon Kinesis con Amazon EMR

Quando analizzi i dati su un flusso Amazon Kinesis attivo, dati i limiti imposti dal timeout e da una durata massima per ogni iterazione, è importante eseguire l'analisi di frequente per raccogliere dettagli periodici dal flusso. Esistono diversi modi per eseguire tali script e query a intervalli regolari; consigliamo l'uso di AWS Data Pipeline per attività ricorrenti come queste. Per ulteriori informazioni, consulta le sezioni relative a [PigActivity di AWS Data Pipeline](#) e [HiveActivity di AWS Data Pipeline](#) nella Guida per gli sviluppatori di AWS Data Pipeline.

S3DistCp (s3-dist-cp)

Apache DistCp è uno strumento open source ideale per copiare grandi quantità di dati. S3DistCp è simile a DistCp, ma è ottimizzato per funzionare con AWS, in particolare Amazon S3. Il comando per S3DistCp in Amazon EMR versione 4.0 e successive è `s3-dist-cp`, che puoi aggiungere come fase in un cluster o nella riga di comando. S3DistCp ti consente di copiare in modo efficiente grandi quantità di dati da Amazon S3 in HDFS, nel quale possono essere elaborati da fasi successive nel cluster Amazon EMR. Puoi inoltre utilizzare S3DistCp per copiare dati tra bucket Amazon S3 o da HDFS ad Amazon S3. S3DistCp è più efficiente e scalabile per la copia parallela di una grande quantità di oggetti tra bucket e tra account AWS.

Per i comandi specifici che dimostrano la flessibilità di S3DistCP in scenari reali, consulta l'articolo [Seven tips for using S3DistCp \(Sette consigli per l'utilizzo di S3DistCp\)](#) nel blog sui Big Data AWS.

Come DistCp, S3DistCp usa MapReduce per copiare in modo distribuito. Condivide la copia, la gestione degli errori, il ripristino e le attività di creazione di report su più server. Per ulteriori informazioni sul progetto open source Apache DistCp, consulta la [DistCp guide \(Guida su DistCp\)](#) nella documentazione di Apache Hadoop.

Se S3DistCp non è in grado di copiare alcuni o tutti i file specificati, la fase del cluster ha esito negativo e restituisce un codice di errore diverso da zero. In questo caso, S3DistCp non elimina i file parzialmente copiati.

Important

S3DistCp non supporta nomi di bucket Amazon S3 contenenti il trattino basso.



S3distCP non supporta la concatenazione per i file Parquet. Usa invece PySpark. Per ulteriori informazioni, consulta la sezione relativa alla [Concatenazione dei file Parquet in Amazon EMR](#).

Per evitare errori di copia quando si utilizza S3DistCP per copiare un singolo file (anziché una directory) da S3 a HDFS, utilizza Amazon EMR versione 5.33.0 o successive o Amazon EMR versione 6.3.0 o successive.

Opzioni S3DistCp

Per quanto simile a DistCp, S3DistCp supporta una serie di opzioni differenti per modificare le modalità di copia e compressione dei dati.

Quando richiami S3DistCp, puoi specificare le opzioni descritte nella tabella seguente. Le opzioni vengono aggiunte alla fase con l'elenco di argomenti. Esempi di argomenti S3DistCp sono illustrati nella seguente tabella.

Opzione	Descrizione	Obbligatorio
<code>--src=LOCATION</code>	<p>Posizione dei dati da copiare. Può essere una posizione HDFS o Amazon S3.</p> <p>Esempio: <code>--src=s3:// DOC-EXAMPLE-BUCKET 1 /logs/j-3GYXXXXXX9I0J/node</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> Important</p> <p>S3DistCp non supporta nomi di bucket Amazon S3 contenenti il trattino basso.</p> </div>	Si
<code>--dest=LOCATION</code>	<p>Destinazione per i dati. Può essere una posizione HDFS o Amazon S3.</p> <p>Esempio: <code>--dest=hdfs:///output</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> Important</p> <p>S3DistCp non supporta nomi di bucket Amazon S3 contenenti il trattino basso.</p> </div>	Si
<code>--srcPattern=PATTERN</code>	<p>Un'espressione regolare che filtra l'operazione di copia per un sottoinsieme dei dati in <code>--src</code>. Se non è specificato né <code>--srcPattern</code> né <code>--groupBy</code>, tutti i dati in <code>--src</code> vengono copiati su <code>--dest</code>.</p> <p>Se l'argomento dell'espressione regolare contiene caratteri speciali quali l'asterisco (*), l'espressione regolare o l'intera stringa <code>--args</code> deve essere racchiusa tra virgolette singole (').</p>	No

Opzione	Descrizione	Obbligatorio
	Esempio: <code>--srcPattern=.*daemons.*-ha doop-.*</code>	
<code>--groupBy=PATTERN</code>	<p>Un'espressione regolare che determina il concatenamento da parte di S3DistCp dei file che soddisfano l'espressione stessa. Ad esempio, è possibile utilizzare questa opzione per riunire in un unico file tutti i file di log scritti in un'ora. Il filename concatenato è il valore trovato dall'espressione regolare per il raggruppamento.</p> <p>Le parentesi indicano il modo in cui i file devono essere raggruppati, con tutti gli elementi che soddisfano l'istruzione tra parentesi riuniti in un singolo file di output. Se l'espressione regolare non include un'istruzione fra parentesi, la fase S3DistCp nel cluster ha esito negativo e viene restituito un errore.</p> <p>Se l'argomento dell'espressione regolare contiene caratteri speciali quali l'asterisco (*), l'espressione regolare o l'intera stringa <code>--args</code> deve essere racchiusa tra virgolette singole (').</p> <p>Se è specificato <code>--groupBy</code>, vengono copiati solo i file che corrispondono al modello specificato. Non è necessario specificare <code>--groupBy</code> e <code>--srcPattern</code> contemporaneamente.</p> <p>Esempio: <code>--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+).*</code></p>	No

Opzione	Descrizione	Obbligatorio
<code>--targetSize=SIZE</code>	<p>La dimensione in mebibyte (MiB) dei file da creare in base all'opzione <code>--groupBy</code> . Questo valore deve essere un numero intero. Se <code>--targetSize</code> è impostato, S3DistCp cerca di rispettare queste dimensioni; di fatto i file copiati possono essere più grandi o più piccoli di questo valore. I processi vengono aggregati in base alle dimensioni del file di dati, pertanto è possibile che la dimensione del file di destinazione corrisponda a quella del file di origine.</p> <p>Se i file concatenati da <code>--groupBy</code> hanno una dimensione maggiore rispetto al valore di <code>--targetSize</code> , vengono suddivisi in file part e denominati in sequenza con un valore numerico aggiunto alla fine. Ad esempio, un file concatenato in <code>myfile.gz</code> sarebbe suddiviso in parti come: <code>myfile0.gz</code> , <code>myfile1.gz</code> ecc.</p> <p>Esempio: <code>--targetSize=2</code></p>	No
<code>--appendToLastFile</code>	<p>Specifica il comportamento di S3DistCp durante la copia da Amazon S3 su HDFS di file già presenti. Aggiunge i nuovi dati ai file esistenti. Se utilizzi <code>--appendToLastFile</code> con <code>--groupBy</code> , i nuovi dati vengono aggiunti ai file che corrispondono agli stessi gruppi. Questa opzione rispetta anche il comportamento <code>--targetSize</code> quando usato con <code>--groupBy</code> .</p>	No

Opzione	Descrizione	Obbligatorio
<pre>--outputCodec=CODEC</pre>	<p>Specifica i codec di compressione da utilizzare per i file copiati. I valori possibili sono <code>gzip</code>, <code>gz</code>, <code>lzo</code>, <code>snappy</code> o <code>none</code>. È possibile utilizzare questa opzione, ad esempio, per convertire i file di input compressi con Gzip in file di output con compressione LZO o per decomprimere i file durante l'operazione di copia. Se scegli un codec di output, il filename viene aggiunto con l'estensione appropriata (ad esempio per <code>gz</code> e <code>gzip</code>, l'estensione è <code>.gz</code>). Se non specifichi un valore per <code>--outputCodec</code>, i file vengono copiati senza apportare modifiche alla compressione.</p> <p>Esempio: <code>--outputCodec=lzo</code></p>	No
<pre>--s3ServerSideEncryption</pre>	<p>Garantisce che i dati di destinazione siano trasferiti tramite il protocollo SSL e automaticamente crittografati in Amazon S3 con una chiave lato servizio AWS. Durante il recupero dei dati con <code>S3DistCp</code>, gli oggetti vengono automaticamente decrittografati. Se tenti di copiare un oggetto non crittografato in un bucket Amazon S3 che richiede la crittografia, l'operazione ha esito negativo. Per maggiori informazioni, consulta Utilizzo della crittografia dei dati.</p> <p>Esempio: <code>--s3ServerSideEncryption</code></p>	No
<pre>--deleteOnSuccess</pre>	<p>Se l'operazione di copia viene completata, questa opzione indica a <code>S3DistCp</code> di eliminare i file copiati dal percorso di origine. Questa funzione è utile se stai copiando file di output, ad esempio i file di log, da una posizione a un'altra come attività pianificata e non vuoi copiare lo stesso file due volte.</p> <p>Esempio: <code>--deleteOnSuccess</code></p>	No

Opzione	Descrizione	Obbligatorio
<code>--disableMultipartUpload</code>	Disattiva l'utilizzo del caricamento in più parti. Esempio: <code>--disableMultipartUpload</code>	No
<code>--multipartUploadChunkSize=SIZE</code>	Le dimensioni, in MiB, di ogni parte in un caricamento in più parti di Amazon S3. S3DistCP utilizza il caricamento in più parti quando copia dati di dimensioni superiori rispetto a <code>multipartUploadChunkSize</code> . Per migliorare le prestazioni del processo, puoi aumentare le dimensioni di ogni parte. La dimensione predefinita è 128 MiB. Esempio: <code>--multipartUploadChunkSize=1000</code>	No
<code>--numberOfFiles</code>	Antepone numeri sequenziali ai file di output. Il conteggio inizia da 0, a meno che non venga specificato un valore diverso tramite <code>--startingIndex</code> . Esempio: <code>--numberOfFiles</code>	No
<code>--startingIndex=INDEX</code>	Utilizzato con <code>--numberOfFiles</code> per specificare il primo numero nella sequenza. Esempio: <code>--startingIndex=1</code>	No
<code>--outputManifest=FILENAME</code>	Crea un file di testo, compresso con Gzip, che contiene un elenco di tutti i file copiati da S3DistCp. Esempio: <code>--outputManifest=manifest-1.gz</code>	No

Opzione	Descrizione	Obbligatorio
<code>--previousManifest=PATH</code>	<p>Legge un file manifest creato durante una precedente chiamata a S3DistCp utilizzando il flag <code>--outputManifest</code> . Quando viene impostato il flag <code>--previousManifest</code> , S3DistCp esclude i file elencati nel manifest dall'operazione di copia. Se viene specificato <code>--outputManifest</code> insieme a <code>--previousManifest</code> , i file elencati nel manifest precedente compaiono anche nel nuovo file manifest, pur non venendo copiati.</p> <p>Esempio: <code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	No
<code>--requirePreviousManifest</code>	<p>Richiede un manifest precedente creato durante una precedente chiamata a S3DistCp. Se impostato su <code>false</code>, non vengono generati errori se non viene specificato un manifest precedente. Il valore predefinito è <code>true</code>.</p>	No
<code>--copyFromManifest</code>	<p>Inverte il comportamento di <code>--previousManifest</code> per indicare a S3DistCp di utilizzare il file manifest specificato come un elenco di file da copiare, invece che da escludere dalla copia.</p> <p>Esempio: <code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	No

Opzione	Descrizione	Obbligatorio
<code>--s3Endpoint=ENDPOINT</code>	<p>Specifica l'endpoint Amazon S3 da utilizzare durante il caricamento di un file. Questa opzione imposta l'endpoint sia per l'origine che per la destinazione. Se non è impostato, l'endpoint predefinito è <code>s3.amazonaws.com</code>. Per un elenco degli endpoint di Amazon S3, consulta la sezione relativa a Regioni ed endpoint.</p> <p>Esempio: <code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	No
<code>--storageClass=CLASS</code>	<p>La classe di archiviazione da utilizzare quando la destinazione è Amazon S3. I valori validi sono STANDARD e REDUCED_REDUNDANCY. Se questa opzione non è specificata, S3DistCp cerca di preservare la classe di storage.</p> <p>Esempio: <code>--storageClass=STANDARD</code></p>	No

Opzione	Descrizione	Obbligatorio
<code>--srcPrefixesFile=PATH</code>	<p>Un file di testo in Amazon S3 (s3://), HDFS (hdfs://) o nel file system locale (file:/) che contiene un elenco di prefissi <code>src</code>, uno per riga.</p> <p>Se <code>srcPrefixesFile</code> viene indicato, <code>S3DistCp</code> non elenca il percorso <code>src</code>. Al contrario, genera un elenco di origine combinando tutti gli elenchi di prefissi specificati in questo file. Per generare percorsi di destinazione, invece dei prefissi verrà utilizzato il percorso relativo rispetto al percorso di <code>src</code>. Se viene specificato anche <code>srcPattern</code>, verrà applicato ai risultati dell'elenco combinato dei prefissi di origine per filtrare ulteriormente l'input. Se viene utilizzato <code>copyFromManifest</code>, gli oggetti nel manifest verranno copiati e <code>srcPrefixesFile</code> verrà ignorato.</p> <p>Esempio: <code>--srcPrefixesFile=PATH</code></p>	No

Oltre alle opzioni sopra riportate, in `S3DistCp` è implementata l'[interfaccia Tool](#), il che significa che sono supportate le opzioni generiche.

Aggiunta di `S3DistCp` come fase in un cluster

È possibile effettuare la chiamata a `S3DistCp` aggiungendolo come fase nel cluster. È possibile aggiungere fasi a un cluster all'avvio o in esecuzione utilizzando la console, la CLI o l'API. I seguenti esempi illustrano l'aggiunta di una fase `S3DistCp` su un cluster in esecuzione. Per ulteriori informazioni sull'aggiunta di fasi a un cluster, consulta [Invio di lavoro a un cluster](#) nella Guida alla gestione di Amazon EMR.

Per aggiungere una fase `S3DistCp` a un cluster in esecuzione utilizzando la AWS CLI

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la [Guida di riferimento ai comandi della AWS CLI](#).

- Per aggiungere una fase a un cluster che chiama S3DistCp, trasferire come argomenti i parametri che specificano in che modo S3DistCp deve eseguire l'operazione di copia.

L'esempio seguente copia il log dei daemon da Amazon S3 a `hdfs:///output`. Nel comando riportato qui di seguito:

- `--cluster-id` specifica il cluster
- `Jar` indica la posizione del file JAR di S3DistCp. Per un esempio di come eseguire un comando su un cluster utilizzando `command-runner.jar`, consulta [Invia una fase JAR personalizzata per l'esecuzione di uno script o di un comando](#).
- `Args` è un elenco separato da virgole delle coppie nome-valore da trasferire a S3DistCp. Per un elenco completo delle opzioni disponibili, consulta [Opzioni S3DistCp](#).

Per aggiungere una fase di copia S3DistCp a un cluster in esecuzione, inserisci quanto segue in un file JSON salvato in Amazon S3 o nel file system locale come `myStep.json` per questo esempio. Sostituisci `j-3GYXXXXXX9I0K` con l'ID del cluster e sostituisci `mybucket` con il nome del bucket Amazon S3.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0K/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 Copia dei file di log da Amazon S3 ad HDFS

Questo esempio illustra anche come copiare in HDFS i file di log archiviati in un bucket Amazon S3 aggiungendo una fase a un cluster in esecuzione. In questo esempio l'opzione `--srcPattern` viene utilizzata per limitare i dati copiati nel log dei daemon.

Per copiare i file di log da Amazon S3 a HDFS utilizzando l'opzione `--srcPattern`, inserisci quanto segue in un file JSON salvato in Amazon S3 o nel file system locale come *myStep.json* per questo esempio. Sostituisci *j-3GYXXXXXXXX9I0K* con l'ID del cluster e sostituisci *mybucket* con il nome del bucket Amazon S3.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

Eliminazione dopo attività S3DistCp non riuscite

Se S3DistCp non riesce a copiare alcuni o tutti i file specificati, la fase del cluster o del comando ha esito negativo e restituisce un codice di errore diverso da zero. In questo caso, S3DistCp non elimina i file parzialmente copiati. È necessario eliminarli manualmente.

I file parzialmente copiati vengono salvati nella directory HDFS tmp nelle sottodirectory con l'identificatore univoco dell'attività S3DistCp. È possibile trovare questo ID nell'output standard dell'attività.

Ad esempio, per un'attività S3DistCp con ID 4b1c37bb-91af-4391-aaf8-46a6067085a6, è possibile connettersi al nodo master del cluster ed eseguire il comando seguente per visualizzare i file di output associati all'attività.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output
```

Il comando restituisce un elenco di file simile al seguente:

```
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
```

È quindi possibile eseguire il seguente comando per eliminare la directory e tutti i relativi contenuti.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

Eseguire comandi e script su un cluster Amazon EMR

In questo argomento viene illustrato come eseguire un comando o uno script come fase del cluster. L'esecuzione di un comando o di uno script come fase è uno dei molti modi che puoi [Inviare lavoro a un cluster](#) ed è utile nelle seguenti situazioni:

- Quando non hai accesso SSH al tuo cluster Amazon EMR
- Quando si desidera eseguire un comando bash o shell per risolvere i problemi del cluster

Puoi eseguire uno script sia quando crei un cluster che quando lo stato del cluster è WAITING. Per eseguire uno script prima dell'inizio dell'elaborazione delle fasi, utilizza invece un'operazione di bootstrap. Per ulteriori informazioni, consulta [Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo](#) nella Guida alla gestione di Amazon EMR.

Amazon EMR fornisce i seguenti strumenti per aiutarti a eseguire script, comandi e altri programmi in cluster. Puoi richiamare entrambi gli strumenti utilizzando la console di gestione Amazon EMR o la AWS CLI.

`command-runner.jar`

Situato sull'AMI Amazon EMR per il tuo cluster. Puoi utilizzare `command-runner.jar` per eseguire comandi sul cluster. Specificare `command-runner.jar` senza utilizzare il suo percorso completo.

`script-runner.jar`

Ospitato su Amazon S3 a `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar` dove `<region>` è la Regione in cui risiede il cluster Amazon EMR. Puoi utilizzare `script-runner.jar` per eseguire script salvati localmente o su Amazon S3 sul cluster. Devi specificare l'URI completo di `script-runner.jar` quando invii un passaggio.

Inviare un passaggio JAR personalizzato per l'esecuzione di uno script o di un comando

I seguenti esempi AWS CLI illustrano alcuni casi d'uso comuni di `command-runner.jar` e `script-runner.jar` su Amazon EMR.

Example : esecuzione di un comando su un cluster utilizzando `command-runner.jar`

Quando si utilizza `command-runner.jar`, si specificano comandi, opzioni e valori nell'elenco degli argomenti del passaggio.

Il seguente esempio AWS CLI invia un passaggio a un cluster in esecuzione che invoca `command-runner.jar`. Il comando specificato nell'elenco Args scarica uno script chiamato *my-script.sh* da Amazon S3 nella home directory dell'utente hadoop. Il comando modifica quindi le autorizzazioni dello script e viene eseguito *my-script.sh*.

Quando si utilizza la AWS CLI, gli elementi nel tuo elenco Args deve essere separato da virgole senza spazi bianchi tra gli elementi dell'elenco. Ad esempio, `Args=[example-command,example-option,"example option value"]` anziché `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 : esecuzione di uno script su un cluster utilizzando `script-runner.jar`

Quando si utilizza `script-runner.jar`, si specifica lo script che si desidera eseguire nell'elenco degli argomenti del passaggio.

Il seguente esempio AWS CLI invia un passaggio a un cluster in esecuzione che invoca `script-runner.jar`. In questo caso, lo script denominato *my-script.sh* è archiviato su Amazon S3. È possibile specificare anche script locali archiviati nel nodo principale del cluster.

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Run a script from S3 with script-  
runner.jar",ActionOnFailure=CONTINUE,Jar=s3://us-west-2.elasticmapreduce/libs/script-  
runner/script-runner.jar,Args=[s3://EXAMPLE-DOC-BUCKET/my-script.sh]
```

Altri modi per utilizzare `command-runner.jar`

È possibile utilizzare anche `command-runner.jar` per inviare lavori a un cluster con strumenti come `spark-submit` o `hadoop-streaming`. Quando si avvia un'applicazione utilizzando `command-runner.jar`, specifici `CUSTOM_JAR` come tipo di passaggio invece di utilizzare un valore come `SPARK`, `STREAMING`, oppure `PIG`. La disponibilità degli strumenti varia a seconda delle applicazioni installate sul cluster.

Il seguente comando di esempio utilizza `command-runner.jar` per inviare un passaggio utilizzando `spark-submit`. L'elenco `Args` specifica `spark-submit` come comando, seguito dall'URI Amazon S3 dell'applicazione Spark `my-app.py` con argomenti e valori.

```
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]
```

La tabella seguente identifica gli strumenti aggiuntivi che è possibile eseguire utilizzando `command-runner.jar`.

Nome dello strumento	Descrizione
<code>hadoop-streaming</code>	Invia un programma di streaming Hadoop. Nella console e in alcuni SDK, questa è una fase di streaming.
<code>hive-script</code>	Esegue uno script Hive. Nella console e negli SDK, questa è una fase Hive.
<code>pig-script</code>	Esegue uno script Pig. Nella console e negli SDK, questa è una fase Pig.
<code>spark-submit</code>	Esegue un'applicazione Spark. Nella console, questa è una fase Spark.
<code>hadoop-lzo</code>	Esegue l'indicizzatore Hadoop LZO su una directory.

Nome dello strumento	Descrizione
s3-dist-cp	Copia distribuita di grandi quantità di dati da Amazon S3 in HDFS. Per ulteriori informazioni, consulta S3DistCp (s3-dist-cp) .

Glossario per AWS

Per la terminologia AWS più recente, consultare il [glossario AWS](#) nella documentazione di riferimento per Glossario AWS.